Demand-Side Management Policy: Mechanisms for Success and Failure

by

Peter Warren

August 2015

PhD thesis submitted in fulfilment of the degree of Doctor of Philosophy in Energy Policy

UCL Energy Institute
University College London (UCL)

Declaration

I, Peter Warren, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

August 2015

Peter Warren

Acknowledgements

This research was made possible by UK EPSRC support, grant number EP/H009612/1. I would like to thank my two supervisors, Dr Mark Barrett and Professor Paul Ekins for their guidance throughout the PhD. I would also like to thank Andrew Smith, Dr David Shipworth, Michelle Shipworth and Professor Helen Roberts for their advice on the development of the methodology. I am similarly grateful to the research participants for giving up their time to take part in the multi-criteria decision-making analysis interviews. I am also grateful to the two examiners, Professor Diana Ürge-Vorsatz and Dr Steven Sorrell, for providing detailed feedback on the thesis. Finally, I give thanks to my family, friends and partner for their encouraging support over the three and a half years of the PhD.

Statement of Copyright

Abstract

Demand-side management (DSM) refers to activities undertaken on the demand-side of energy meters that seek to meet energy policy objectives. The policy side of DSM has received limited attention in the academic literature, with previous research focussing on technological trials, utility programmes and modelling studies of the potential of DSM. Within the field of DSM policy evaluation, much of the work has concentrated on policy impacts rather than policy mechanisms. The thesis contributes to filling this research gap by determining the key factors for success and failure for various DSM policies and countries (and sub-national states).

A global systematic review of the DSM policy evaluation evidence was conducted. The method included the critical appraisal of the quality of the evidence base and the final sample included 119 high quality documents (covering 690 evaluations) from 35 databases, covering 30 countries, 36 sub-national states, and 21 individual DSM policies and policy packages. A technique was developed to combine factor frequency and weighting analyses in order to establish the success and failure factors that were both frequent and highly weighted for given DSM policies and countries/states.

Overall, across policies and countries/states, regulatory frameworks and appropriate incentives are the most crucial success factors, and a lack of monitoring and technical issues (primarily programme management issues) are the most crucial failure factors. California, China, the UK and the USA have experienced the greatest success with DSM policies, each having successfully implemented and evaluated 9-10 policies. Utility obligations, performance standards and alternative utility business models have been the most successful policies overall, whilst labelling, information campaigns, and loans and subsidies have been the least successful. However, all policies show examples of both success and failure in specific contexts and the research has identified which key factors cause various demand-side policies to succeed or fail.

Table of Contents

List of Figures

List of Tables

List of Equations

List of Appendix Figures

List of Abbreviations

DSM: Demand-Side Management

DR: Demand Response (also DSR – Demand-Side Response)

MCDM: Multi-Criteria Decision-Making Analysis

IPBDR: Incentive Payment-Based Demand Response

PBDR: Price-Based Demand Response

MT: Market Transformations

IR: Infrastructure Rollouts

UO: Utility Obligations

LB: Labelling

PS: Performance Standards

L&S: Loans and Subsidies

UBM: Utility Business Models

R&D: Research and Development

IC: Information Campaigns

VP: Voluntary Programmes

FW_{pf} = Frequency-Weighting combined analysis

PS_{p} = Policy Success weighting

PSF_{pf} = Policy Success Factor Frequency

PSW_{pf} = Policy Success Factor Weighting

$FW_{pf\%}$ = Frequency-Weighting combined analysis percentage

FW_{pf} = Frequency-Weighting combined analysis

FW_{pfmax} = Theoretical Maximum combined analysis

P_{p} = Policy Frequency

PS_{pmax} = Theoretical Maximum Policy Success Weighting

1 Chapter 1: Introduction

1.1 Energy Policy Objectives

Environmental and energy security issues are increasingly moving to the forefront of the political agenda. Energy production and consumption is widely regarded as a key contributor to anthropogenic climate change, and the International Energy Agency (IEA) estimates that around 70% of world energy production is produced through the burning of fossil fuels, primarily coal (42%) and gas (21%), and energy accounts for 40% of anthropogenic carbon dioxide and other greenhouse gas emissions (CO_{2e}) (IEA, 2012). The demand for energy is increasing due to population growth, particularly in emerging economies, and the continued growth of gadgets, electronics, technologies and appliances in society (IEA, 2009; Cabeza *et al*., 2014).

Balancing energy supply and demand has been a complex challenge in many countries, with reserve capacity margins of around 20% commonly used to deal with peak demands (Anderson, 2006), such as when people turn their kettles on after a popular television programme or turn their heating systems on during a particularly cold winter night (Bunn and Seigal, 1983). However, with flexible generation plants powered by fossil fuels, matching supply with demand has been effectively administrated in most countries. Traditionally, energy utilities have invested in expanding their fossil fuel capacity base to deal with long-term increases in energy demand (Torriti *et al*., 2010). However, with growing awareness regarding the contribution of fossil fuel generation to climate change, energy utilities are coming under political pressure to diversify their fuel mixes to lower carbon alternatives, such as wind and solar power. Nevertheless, wind power suffers from variable power production due to wind speed variations and solar power output is dependent on the availability of sufficient sunlight, thus causing new challenges in matching supply and demand (Torriti *et al*., 2010). Developing lower carbon options that can meet peak and variable demands is one of the crucial energy policy challenges of the 21st century.

Furthermore, a growing number of countries, particularly in Europe, are becoming more dependent on fuel imports, such as coal, oil and gas, than domestic supplies. In some cases the imports are sourced mainly from specific regions, such as Europe's dependence on Russian gas and Middle Eastern oil (Bahgat, 2006). The dominance of fossil fuel energy resources has given countries in these exporting regions increasing geopolitical power as energy moves up the political agenda (Bahgat, 2006). Hence, growing energy demands, the political drive to move to lower carbon energy sources, and the growing dependence on fuel imports, have resulted in renewed debates regarding the security of energy supply. The European Commission defines 'energy security' as:

"The ability to ensure that future essential energy needs can be met, both by means of adequate domestic resources worked under economically acceptable conditions or maintained as strategic reserves, and by calling upon accessible and stable external sources supplemented where appropriate by strategic stocks." (EUROGULF, cited in Bahgat, 2006)

The use of the word 'resources' in the definition is important, as it indirectly includes non-traditional resources in addition to traditional resources (such as coal, oil and gas power plants). Nevertheless, the definition does not explicitly state this and as such the European Commission should adopt a clearer definition. Meeting potentially competing policy objectives, such as energy security and carbon emissions reduction, is a current challenge for many governments around the world. Proposed solutions to this challenge include building new low(er) carbon capacity, increasing interconnections with other countries, developing energy storage technologies and utilising demand-side management (Barrett, 2006). These options are complementary and Droste-Franke *et al*. (2012) argue that they will all be important in the future.

In Europe, political pressure is mounting on energy utilities to invest in new capacity that is low(er) carbon. Nuclear power, which uses uranium as a fuel source rather than a carbon-based fuel source, has been pursued in a number of countries as an alternative to fossil fuel-based power production. However, following the Fukushima-Daiichi disaster in Japan in March 2011, where an earthquake-triggered tsunami devastated the Fukushima nuclear power plant, many governments have started to question their nuclear policies (Wittneben, 2012), such as in Germany. There were similar reactions following previous accidents at Chernobyl in 1986 (in the former Soviet Union) and Three Mile Island in 1979 (in the USA). Furthermore, nuclear has been used as base load for technical and economic reasons, such as its inflexible operational nature (Verbruggen, 2008). Many of the alternatives to nuclear, wind and solar power are underdeveloped and at the demonstration stages, such as wave and tidal power, and carbon capture and storage (CCS) technologies. The former uses the power of waves or the tides to drive turbines and generate electricity, whereas the latter captures the carbon dioxide emissions from fossil fuel-based power plants and stores them underground. Many of these options are currently expensive due to being in the early stages of commercial maturity. A more commercially mature alternative is bio-energy. Some countries, such as the UK, have begun converting coal plants into biomass-burning plants (for example, the Drax power station). However, due to uncertainty in the sustainability of bioenergy, some have argued that a precautionary approach to its development should be taken (McDowall *et al*., 2012; Thornley *et al*., 2009).

Building new capacity as back-up power is costly as the power plants are used infrequently. Alternatively, there is a growing interest in the role that interconnections can play, particularly in the common European market. Interconnections refer to the cross-border transmission of electricity along high voltage direct current (HVDC) power lines between countries, though this requires the right infrastructure and regulatory transaction processes to be in place (Galarraga *et al*., 2011, p. 5). For example, the UK currently has interconnections with France, the Netherlands, Northern Ireland and Ireland with a combined capacity of 4 GW, and it is currently developing a 1 GW link to Belgium, a 1 GW second link to France and a 1.4 GW link to Norway (National Grid, 2015). The details of the UK interconnections that are existing, under development and proposed are shown overleaf (taken from National Grid, 2015).

Existing:

- IFA 1 to France (2 GW capacity, 70 km long)
- BritNed to Netherlands (1 GW capacity, 260 km long)
- Moyle to Northern Ireland (500 MW capacity, 64 km long)
- East-West to Ireland (500 MW capacity, 261 km long)

Under Development:

- Nemo Link Limited to Belgium (1 GW capacity, 130 km long, operational by 2018)
- IFA 2 to France (1 GW capacity, 140 km long, operational by 2020)
- NSN to Norway (1.4 GW capacity, >700 km long, operational by 2021)

Proposed:

- Viking Link to Denmark (1-1.4 GW, 600-700 km long, feasibility stage, operational by 2020)
- Ice Link to Iceland (0.8-1.2 GW, 1,000 km long, feasibility stage)

The UK plans to use interconnections not only to contribute to meeting energy security needs but also to tap into the renewable energy capacity of other countries in order to meet domestic renewable energy targets (for example, importing renewable electricity from Ireland and Denmark's wind farms, and potentially from Iceland's geothermal plants and Norway's hydro-electric plants). Nevertheless, unless interconnections are more far reaching geographically, they may make little difference to countries experiencing the same weather patterns if wind and solar are pursued as major power sources (UK Parliament, 2011).

Energy storage is likely to play an important role in the future but many storage technologies are not currently commercially mature. Pumped hydro storage is one of the few commercially available and widely used technologies, but it has geographical limitations in the extent of its development (Deane *et al.*, 2010). Most countries with the necessary mountainous and river terrain have already developed their most suitable sites. Pumped hydro refers to a hydro-electric plant with two reservoirs at different elevations. During times of low demand and cheaper electricity prices (off-peak periods), water is pumped from the lower reservoir to the higher reservoir, and during times of peak demand when prices are high, water flows under gravity from the higher reservoir to the lower reservoir to drive turbines to generate electricity, which is then fed into the electricity grid (Deane *et al*., 2010). The geographically distributed nature of variable renewable sources may prevent certain energy storage systems from being practicably installed (Beaudin *et al*., 2010), though some have shown promise, such as batteries connected to wind turbines (Divya and Østergaard, 2009). Other large-scale storage options include flywheels (a rotating mechanical device used to store rotational energy) and compressed air energy storage (the compression and storage of air in large repositories, such as underground salt caverns). However, these technologies have not been economically proven to date. In contrast, smaller scale storage options such as electric vehicle batteries (charging batteries at night when demand is low and releasing electricity to the grid when demand is high) and large thermal storage tanks (storing hot water in highly insulated tanks) (Evans *et al*. 2012) are currently commercially available.

Many of the proposed solutions to current policy objectives have developed from the traditional approach of matching supply with demand. Demand-side management (DSM) aims to reverse this thinking by looking at how to match demand with the available supply. DSM complements the other solutions and engages consumers in a market that has generally been ‘invisible’ to them (Darby 2006, p. 3), and meeting climate change and energy security policy objectives is likely to require changes in behaviour in addition to cleaner technologies (Chatterton, 2011). DSM involves activities, technologies and programmes on the demand-side of energy meters, such as energy efficiency (e.g. installing insulation in buildings), demand response (e.g. organisations being paid to reduce consumption during peak times), and on-site generation and storage (e.g. solar photovoltaics on buildings). This is discussed further in chapter two.

1.2 Thesis Overview

DSM is not a new concept and there is much research on the topic dating back to the energy crises of the 1970s. However, a literature review of 389 academic, industrial and policy documents highlights that DSM policy is an understudied area within the field, with the majority of previous studies focussing on DSM trials, utility programmes and modelling the future potential of DSM. The previous research that has been conducted on the policy side of DSM has concentrated on the quantitative impacts of implemented policies, particularly in terms of energy and carbon savings and cost-benefit analyses. However, going beyond impacts to look at the mechanisms behind how and why policies performed as they did is a much under-researched area. Thus, the thesis has the following research aim and research questions:

Research aim:

1. To determine the mechanisms behind DSM policy success and failure

Research questions:

1. What DSM policies have been implemented around the world with high quality documented evaluations?
2. How and why do DSM policies succeed or fail, and what policies have been successful?

The literature review identified that the quality of the evidence base for DSM policy evaluation has not been established and this is the justification for research question one. The research question aims to map out the countries that have implemented DSM policies and produced high-quality evaluations of those policies. Research question two forms the central part of the thesis and aims to determine the key factors that cause different types of DSM policy to succeed or fail. The research question also seeks to identify how successful different types of demand-side policies have been around the world.

The thesis is split into six chapters. Chapter two discusses DSM policy theory. Firstly, it gives background to DSM in terms of the contested definitions of DSM, the benefits and challenges of DSM, the history of DSM in policy since the energy crises of the 1970s, and discusses current international experiences and the role of DSM in the future smart(er) grid. Secondly, the chapter discusses policy theory from the political science literature before examining the theory and practice of DSM policy evaluation.

Chapter three focuses on research design. Firstly, it details the research focus and the methodological approach underpinning the thesis. Secondly, the chapter outlines and justifies the methods and processes for data collection and analysis in order to answer the research questions.

Chapter four answers research question one on DSM policy implementation and evaluation. Firstly, it gives an overview of the data collection process for the primary and secondary methods. Secondly, the chapter gives overall statistics from the data collection before discussing the main spatial and temporal patterns for DSM policy implementation and evaluation. The chapter finishes with the main conclusions for research question one.

Chapter five answers research question two on DSM policy mechanisms. Firstly, it discusses and justifies the definitions for policy success and failure. Secondly, the chapter details the results for success and failure in terms of the key overall success and failure factors, statistical associations between factors, the key success and failure factors by DSM policy, and the key success and failure factors by country/state. Thirdly, it identifies the countries that have experienced success with various DSM policies. The chapter finishes with the main conclusions for research question two.

Chapter six provides the main conclusions to the research. Firstly, it discusses the key findings for each research question and identifies the original contributions to knowledge in terms of conceptual, methodological and empirical contributions. Secondly, the chapter outlines the key policy recommendations of the thesis and identifies areas for further research.

2 Chapter 2: DSM Policy Theory

2.1 DSM Background

2.1.1 Contested Definitions

In broad terms, demand-side management (DSM) refers to actions undertaken on the demand side (i.e. customer side) of energy meters (Gellings and Chamberlin, 1993, p. 2). Clark Gellings at the US-based Electric Power Research Institute (EPRI) first coined the term 'demand-side management' in 1984 (Gellings, 1985). DSM programmes focus on the management of electricity demand and/or non-electric based heating and transport, and in the past they have been implemented across different sectors, such as the residential, commercial, public, industrial, transport and agricultural sectors. However, as justified in this chapter and in chapter three, the research focuses on the building-related sectors (residential, commercial, public) and the industrial sector, and on policies that primarily focus on electricity, though policies that cover both electricity and non-electric-based heating (and related measures like insulation) are included due to the interaction of energy demands in buildings. Although residential energy consumption (particularly for heating and cooling) has stagnated in developed countries, commercial energy consumption (primarily for heating and cooling) is increasing in developed and developing countries (Urge-Vorsatz *et al*., 2015). As such, the buildings-related sectors are interesting to examine, in addition to the industrial sector where much of the DSM focus has been in the past.

A literature review of 389 documents (primarily journal papers, books, reports, government documents, interviews, and audiovisual material) that have been published since the energy crises of the 1970s highlighted that definitions of DSM have varied over time in what they include or exclude. Some publications include the management of electricity demand but not other forms of energy demand (e.g. Prüggler *et al*., 2011), others use the definition synonymously with that of the smart(er) grid (discussed in sub-section 2.2.3) (e.g. Davito *et al*., 2010), some refer to DSM as measures that reduce energy demand at peak times (e.g. Ofgem, 2010), while others use a similar definition but also include the response of consumers to price changes and the shifting of load to off-peak times (e.g. Strbac, 2008). Micro-generation is included in some definitions (e.g. Eissa, 2011) and some include or exclude energy efficiency measures (e.g. Sioshansi and Vojdani, 2001). Micro-generation is defined in the UK *Energy Act 2004* as technologies that produce heat and/or electricity from a low carbon source and are <100 kW in size. The International Energy Agency (IEA) defines energy efficiency as delivering more services for the same energy input or delivering the same services for less energy input (2014). Both definitions are established in the literature unlike those for the broader term of DSM. Gellings and Chamberlin (1993) proposed a holistic definition of DSM:

"DSM activities are those which involve actions on the demand (i.e. customer) side of the electric meter, either directly or indirectly stimulated by the utility. These activities include those commonly called load management, strategic conservation, electrification, strategic growth or deliberately increased market share." (pp. 2-3)

Here, Gellings and Chamberlin (1993) include energy efficiency in addition to load management and put the focus on how the load *shape* might vary. This is discussed further in sub-section 2.1.2. Although the definition is arguably more comprehensive, the main limitations are that it only focuses on electricity (and not non-electric based heating and transport) and activities stimulated by electric utilities (rather than national or local government policy, third parties or consumers directly). Gellings and Chamberlin (1993) argue that DSM tries to encourage utilities to put demand-side measures on an equal level with supply-side options (pp. 3-4). It also aims to actively engage consumers in the management of their energy use and how they could save money through making their consumption more 'visible' and important to them (Stromback *et al*., 2011, p. 13). If overall energy demand is reduced (rather than load shifting), DSM can reduce carbon dioxide emissions equivalent (CO_{2e} – a metric that equates all emitted gases to the global warming potential of carbon dioxide) if it offsets energy produced from fossil fuel generation.

The primary issue with Gellings and Chamberlin (1993)'s definition is that it does not directly relate to current policy objectives to reduce CO_{2e} emissions and consumer energy bills in addition to ensuring energy security. Although energy policy objectives vary from one country to another, these objectives are becoming increasingly prominent in Europe, North America and east-Asia, which are the three continents where much of the past and present DSM experience is located. As such, these continents are discussed throughout the thesis in more depth than other continents. Energy policy objectives were discussed in chapter one and are examined in chapter four in relation to DSM policy.

Almost twenty years after Gellings and Chamberlin (1993)'s definition, Eissa (2011) argues that the overall goal of DSM should now be to reduce overall energy demand and shift patterns of consumption to help smooth demand. Thus, the definition incorporates energy conservation (an overall reduction in energy use), energy efficiency and demand response (the response of consumers to incentive payments or price changes – Albadi and El-Saadany, 2008), and is more in-line with current policy objectives to minimise the costs of meeting environmental targets and ensuring energy security. However, Eissa (2011)'s definition excludes some aspects of load management that were also used by utilities in the 1970s-1990s, such as strategic load growth, where utilities deliberately increase loads in times of excess capacity (Gellings and Chamberlin, 1993). This could become more common under conditions of surplus wind power where there are limited storage capabilities and limited international demand through interconnections. Despite this, the key strength of Eissa (2011)'s definition, which is more comprehensive and arguably more relevant to the 21st century than Gellings and Chamberlin (1993)'s definition, is that it includes the full range of demand-side activities and implementers (governments and third parties in addition to utilities).

Furthermore, from a technological point of view, Eissa (2011)'s definition also includes micro-generation in addition to energy efficiency, energy conservation and demand response, which was not included in a number of the definitions from the mid-1980s to the mid-2000s. Micro-generation from on-site technologies, such as solar photovoltaic panels, small wind turbines, and micro-hydro plants, is included with respect to a reduction in demand for electricity from national grids and public networks. Micro-generation can also refer to the small-scale generation of heat from on-site technologies, such as solar thermal panels, heat pumps and biomass boilers, and is included with respect to a reduction in demand for heat through gas grids, district heating, or industrial waste heat networks. Eissa (2011)'s definition is discussed further below.

In the UK, the term 'D3' has developed in recent years to refer to 'demand reduction', 'demand response' (DR – also known as demand-side response (DSR)) and 'distributed generation' (DECC, 2014a). This is a useful term to refer to all demand-side activities. This generally fits with the definition of DSM, which includes demand reduction through energy efficiency and energy conservation, demand response and on-site generation. However, it is the latter group where clarification is needed and the boundaries are drawn. The reason why 'D3' does not fully equate to DSM is due to the definition of 'distributed energy'. DECC (2014a) defines this group as the use of low carbon electricity and heat located on-site or within the local distribution network. DSM does not include technologies that are not on-site to buildings. Distributed energy primarily refers to small-scale generation plants that are connected directly to distribution networks rather than transmission networks or directly connected to energy meters within the buildings of consumers. Although the plants are usually medium-scale and between 100 kW-2 MW in size rather than <100 kW as per the definition of micro-generation, DSM can refer to plants >100 kW if they are on-site to buildings. For example, it is common to have medium-scale back-up generators (usually diesel generators) in large industrial consumers.

Finally, as per the 'D3' definition, distributed energy also includes district heating, industrial waste heat networks, off-site anaerobic digestion plants, off-site medium-scale heat pumps and off-site medium-scale combined heat and power (CHP) systems (the production of both heat and electricity in one process – CHPA, 2014). These technologies and processes do not come under the definition of DSM, as they are not located on-site to buildings. However, micro-CHP (<100 kW) is included within the definition of DSM, as it is a form of micro-generation technology. CHP schemes that are 100 kW to 2 MW in size are included in the definition if they are located on-site, such as within a large industrial consumer. In summary, 'D3' is synonymous with DSM except distributed energy that is not located on-site to buildings.

From the 389 documents consulted in the literature review, twenty-two documents explicitly state definitions of DSM (rather than simply describing it or discussing relevant technologies). A large number of them have been published in the 2000s or in the first half of the 2010s. The differences between them highlight the variations in what is considered to constitute DSM, but it is a crucial part of justifying the need for a new definition that combines the definitions into a more holistic one. A selection of the other definitions is listed below to highlight the variations:

"Demand Side Management [is] [t]he planning, implementation, and monitoring of utility activities designed to encourage customers to modify patterns of electricity usage, including the timing and level of electricity demand. Demand-Side Management (DSM) covers the complete range of load-shape objectives, including strategic conservation and load management, as well as strategic load growth." (Kerr *et al*., 2011)

Kerr *et al*. (2011)'s definition draws parallels to the historical definitions of the 1980s and 1990s, as developed by Gellings (such as the previously discussed Gellings, 1984 and Gelling and Chamberlin, 1993). Kerr *et al*. (2011) argue that DSM does not just refer to programmes that reduce overall energy demand but those that smooth out the supply-demand balance through load management and increase consumption through strategic load growth.

"The very broad definition…includes both modifications of electricity consumption by consumers in response to price and the implementation of more energy efficient technologies." (Greening, 2010)

Greening (2010)'s definition is narrower than that of Gellings and Chamberlin (1993) or Kerr *et al*. (2011), as it only includes demand response to price changes and energy efficiency measures. Other publications include demand response to incentive payments as well as price changes (such as Albadi and El-Saadany, 2008), while others include micro-generation, as previously discussed in relation to Eissa (2011). A further key difference is that Greening (2010) concentrates more on consumer actions rather than those of utilities, as is the case with Gellings and Chamberlin (1993) and Kerr *et al*. (2011). Nevertheless, neither definition includes third parties, local governments or national governments as the implementers of DSM activities. The definition shown overleaf aligns itself more with Greening (2010).

"…many initiatives have been implemented to change consumers' behaviour towards a more efficient one. These initiatives are referred to as demand side management (DSM)." (Didden and D'haeseleer, 2003)

Didden and D'haeseleer (2003) focus more on behavioural practices and actions that result from DSM programmes rather than specific technological or economic tools, and thus take a more sociological definition of DSM. In contrast, Strbac (2008) takes a more technical definition:

"DSM as shifting load from peak to off-peak periods…redistributes the load but does not necessarily reduce the total energy consumed by the device (appliance)." (Strbac, 2008)

The definition is narrower than that of Greening (2010) as it only includes the shifting of loads from peak to off-peak periods, highlighting that DSM is used more for the purposes of supply-demand balancing rather than reducing overall energy consumption. A criticism of this definition is that it not only focuses just on demand response (excluding energy efficiency, energy conservation and micro-generation), but it focuses only on one aspect of demand response – load shifting from peak to off-peak periods.

"Although the term DSM was originally defined to cover a broad range of programs, the focus of efforts fell into two general areas: conservation (i.e., reducing energy use across all hours) and load management (i.e., reducing peak demand)." (Sioshansi and Vojdani, 2001)

Sioshansi and Vojdani (2001), in contrast to Strbac (2008), Kerr *et al*. (2011), and Gellings and Chamberlin (1993), look more at the use of DSM technologies to reduce overall energy consumption and to reduce consumption during peak times rather than including strategic load growth or facilitating behavioural changes. This definition is one of the more holistic definitions discussed in this sub-section, as it does not specify technological categories and processes (such as energy efficiency, energy conservation, demand response and micro-generation) in order to argue that they are all relevant. A further strength of the definition is that it does not specify DSM implementers (such as utilities, third parties, local governments, national governments and consumers directly). Despite this, the main criticism is that it does not link directly to current policy objectives for energy security, CO_{2e} emissions reductions and reducing consumer energy bills.

The discussions so far have shown the evolution of the definition of DSM over time from the 1980s to the 2010s. The main conclusion is that in the 1980s-1990s, DSM referred primarily to load management undertaken by utilities, whereas in the 2000s-2010s, DSM refers to a much broader range of activities and implementers (as discussed below). Eissa (2011)'s definition is the most recent in the evolution of what constitutes DSM and is arguably the most comprehensive to date in terms of its relevance to the 21st century:

"Load management is the process of scheduling the loads to reduce the electric energy consumption and/or the maximum demand…such as [through] load shedding and restoring, load shifting, installing energy-efficient processes and equipment, energy storage devices, co-generation and non-conventional sources of energy, and reactive power control…Demand Response is a subset of the broader category of end-use customer energy solutions known as Demand-Side Management (DSM). In addition to Demand Response, DSM includes energy efficiency programs." (Eissa, 2011)

The definition is thorough in defining the boundaries of DSM. It covers a spectrum of load shapes (load shedding and shifting), technologies and processes (energy efficiency, demand response, and small-scale generation), and defines how some terms, such as demand response, energy efficiency and demand-side management, interact with each other. The definition acknowledges that DSM refers to a broad category of end-use consumer solutions. This links to the discussion of distributed energy in the 'D3' definition as it identifies that co-generation (another term for CHP) comes under this category but straddles the boundaries between distributed energy that is DSM and distributed energy that is not DSM. A further key strength of this definition, which has failed to be acknowledged in previous definitions, is the role of storage. Arguably, this is due to the improvements in small-scale on-site storage technologies over time. However, in the residential sector, hot water storage tanks have been available in some countries for a number of decades (for example, they have been available in the UK long before the term 'DSM' was first coined in the 1980s). Nevertheless, they have not formally been acknowledged in definitions to date.

Despite the key strengths of the most recent and holistic of the definitions for DSM, Eissa (2011)'s definition still does not directly link to current policy objectives, nor are there additional explanations (outside of that stated above) on the boundaries of the technologies and implementers. Thus, the evolution of the definition should continue with a new definition that extends that of Eissa (2011) to fill in these gaps.

2.1.2 Proposed Definition

The thesis proposes the following definition for DSM to comprehensively collate all of the definitions published since the 1980s, and to link them to current policy objectives whilst establishing more concrete definitional boundaries:

"Demand-side management (DSM) refers to technologies, actions and programmes on the demand-side of energy meters that seek to manage or decrease energy consumption, in order to reduce total energy system expenditures or contribute to the achievement of policy objectives such as emissions reduction or balancing supply and demand." (Warren, 2014a)

The above, published definition is the concise version of the following, full definition:

"Demand-side management (DSM) refers to technologies, actions and programmes on the demand-side of energy meters, as implemented by governments, utilities, third parties or consumers, to manage or decrease energy consumption through energy efficiency, energy conservation, demand response or on-site generation and storage, in order to reduce total energy system expenditures or to contribute to the achievement of policy objectives, such as emissions reduction, balancing supply and demand or reducing consumer energy bills."

The strengths of the proposed definition are that it includes both technological and behavioural activities, it uses the phrase 'energy meters' rather than 'electric meters' so that (non-electric-based) heat is included in addition to electricity, it lists the common implementers of DSM in addition to utilities (governments, third parties and consumers directly), it uses the broad phrase "manage energy consumption" to include the full range of load shapes from load shifting to load growth (which could be useful in the future to offset surplus wind power in the absence of storage capabilities) but emphasises energy conservation and energy efficiency through the phrase “decrease energy consumption”, it covers the full range of technologies and processes (energy efficiency, energy conservation, demand response, on-site generation and on-site storage), and it directly links to current policy objectives that go beyond just energy security as the primary driver (though it is still an important driver for DSM, as discussed in chapter four) to also include emissions reduction and reducing consumer energy bills. The definition is visualised in figure 1, which also includes additional findings from the literature review on DSM policy. The diagram is split into three parts: DSM categories, DSM implementers and DSM policies.

For DSM categories, figure 1 conveys that DSM includes energy efficiency, energy conservation, price-based demand response, incentive payment-based demand response, on-site generation and on-site storage. For DSM implementers, the proposed definition highlighted that DSM can be undertaken directly by national or local governments, consumers, third parties or utilities. This can also be extended to include system operators, which operate and own the transmission and distribution networks in liberalised markets (the number of system operators varies by country). Additionally, aggregators can also play an important role in implementing DSM. However, in both cases, there is a much stronger focus on demand response than other categories of DSM. For example, system operators may set up contracts directly with large consumers to reduce consumption during peak times (a form of incentive payment-based demand response). Aggregators collate and aggregate smaller load reductions from a number of consumers, which are then offered to a system operator as a combined load reduction in order to participate in balancing, reserve, or capacity markets.

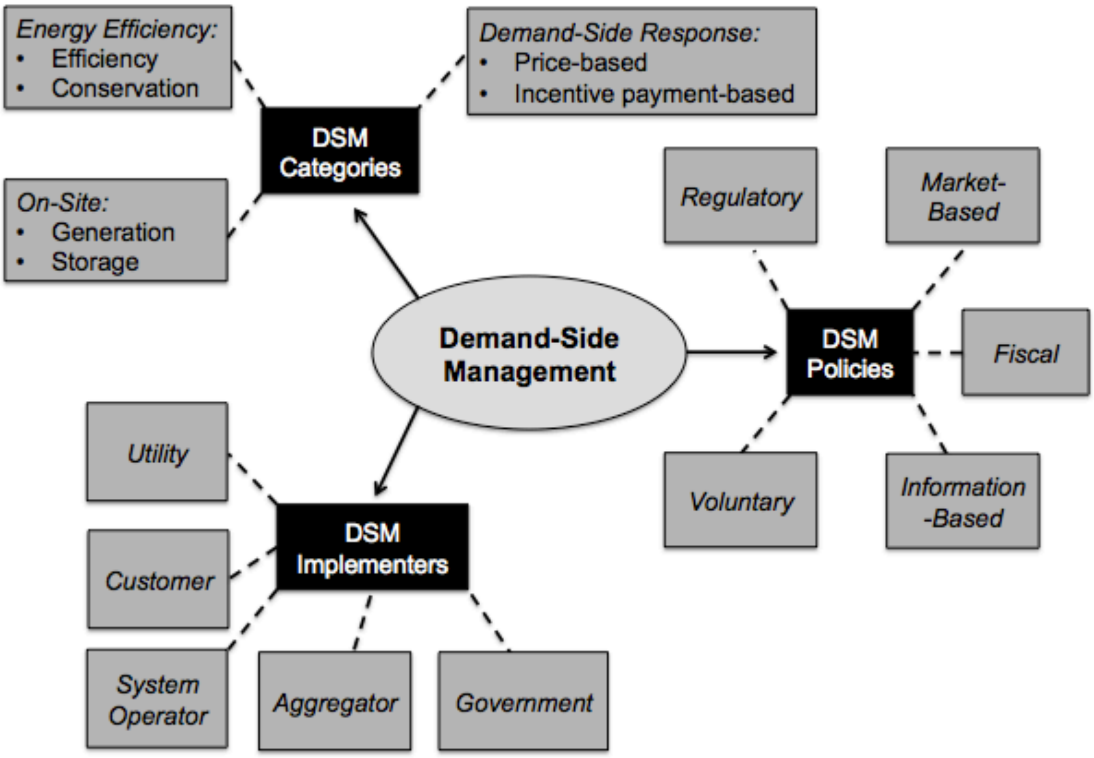


Figure 1: defining the boundaries of demand-side management (DSM)

For DSM policies, DSM can be categorised into regulatory, market-based, fiscal, information-based or voluntary policies. DSM policy is the focus of this research and is detailed in sub-section 2.3.2.

Gellings and Chamberlin (1993) refer to energy efficiency and energy conservation synonymously but they are in fact two different (though related) terms. Energy efficiency is the ratio of the useful output of a process to the energy input into a process (Patterson, 1996). Thus, improving energy efficiency reduces the amount of energy needed to perform the same function or to produce the same services, such as heating, lighting and cooling. However, it is possible that the reduced energy costs from the resulting energy savings could be channelled into producing more of the same services or increasing other energy-consuming activities without increasing overall expenditures (Sorrell *et al*., 2009; Sorrell, 2014). This has been termed the 'rebound effect' (Sorrell, 2007a, pp. v-vi). An example of the rebound effect in the industrial sector is the increased output of manufactured goods following the installation of more energy efficient equipment (Barker *et al*., 2007). In the residential sector, the thermostat controlling the internal temperature may be increased to higher settings to reach greater comfort levels following the installation of energy efficient measures (Sorrell, 2007a, p. 7), such as loft insulation, cavity wall insulation, draught proofing and double glazing. Rebound effects are categorised into direct rebound, where consumption is increased in the same activity, and indirect rebound, where consumption is increased in another activity (Sorrell, 2007a, pp. v-vi; Chitnis *et al*., 2013; Sorrell *et al*., 2009; Druckman *et al*., 2011). In contrast, energy conservation aims to reduce overall energy demand (Davito *et al*., 2010). Thus, if consumption is simply shifted to another time of the day, week, month or year, then it is categorised as demand response rather than energy conservation. Rebound effects can be challenging to measure and Sorrell *et al*. (2009) give a comprehensive review of different approaches. The review argues that studies to date are highly diverse in terms of the definitions, methodological approaches and data sources used.

Figure 2 summarises Gellings and Chamberlin (1993)'s definition of DSM. Although it focuses primarily on load shapes through load management, it still provides a useful visualisation for understanding what DSM is.

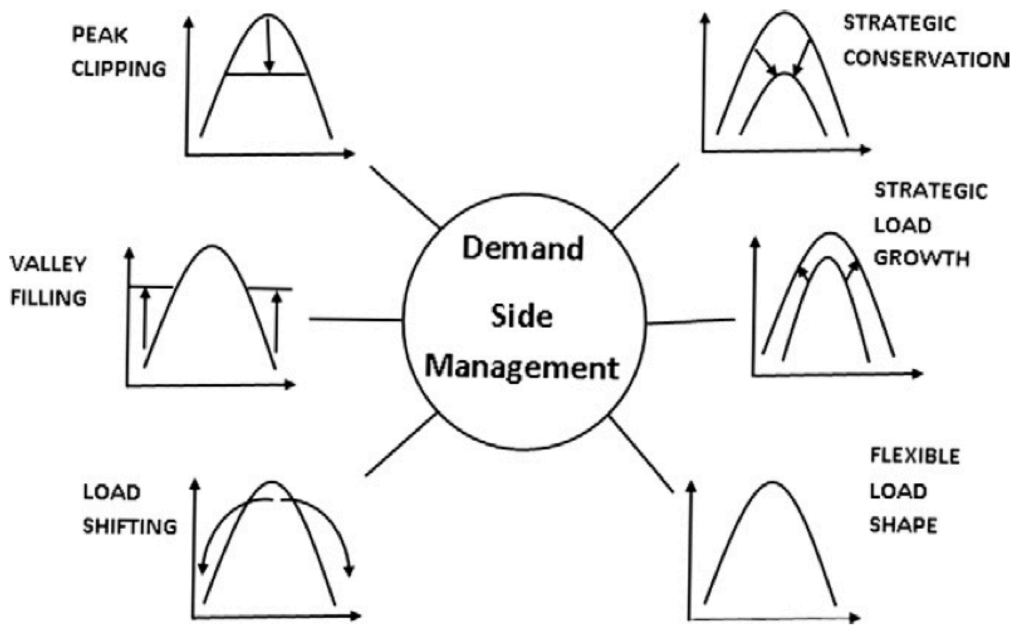


Figure 2: the six main types of load shapes in demand-side management as defined by Gellings and Chamberlin (1993)

Peak clipping refers to the reduction of the energy system peak demand (without shifting consumption to other periods); valley filling refers to building off-peak loads (without necessarily reducing the peak demand); load shifting refers to moving consumption from peak times to other periods or permanently to off-peak periods; strategic conservation refers to an overall reduction in loads; strategic load growth refers to the increase of load in any period (when there is a surplus of generation output); and flexible load shape refers to the customer being presented with options as to the variations in quality of service that they are willing to accept in exchange for various incentives (Gellings and Chamberlin, 1993).

Energy efficiency and energy conservation measures primarily fall under the strategic conservation category. Demand response fits into the other five categories depending on the type of demand response (price-based demand response or incentive payment-based demand response), which are discussed in sub-section 2.3.2. On-site back-up generation and storage have historically been used for smoothing the load curve to reduce peaks and to ensure energy security, such as through diesel generators in industry and hot water storage tanks in houses (Prema *et al*., 2014; Arteconi *et al*., 2012).

From the examination of definitions since the 1980s, it is clear that there has been a long history of DSM implementation. The benefits of DSM are discussed in sub-section 2.1.3 and the current key challenges for DSM are debated in sub-section 2.1.4.

2.1.3 Benefits

In the energy system, there has been much academic debate over the future role of DSM. Many of the arguments are in broad agreement that DSM is a useful and complementary solution to current energy policy challenges (alongside interconnections, large-scale storage, and increased low carbon capacity, as discussed in chapter one), but that a number of challenges remain that need to be overcome in order to increase the contribution of the demand-side. It is clear from the literature review that there are numerous benefits to the implementation of DSM. For society, Gillingham *et al*. (2009) argue that it is a cost-effective way to contribute to overcoming energy security issues, and Pachauri *et al*. (2012) highlight that DSM (particularly energy efficiency) can contribute to addressing energy access issues. For Governments, DSM reduces the risk of ‘black-outs’ (power supply disruptions), which can have important political repercussions during energy crises, and can help to reduce CO_{2e} emissions (Cooke, 2011) in order to meet policy targets. Furthermore, DSM can be an effective way to facilitate behaviour change by helping consumers to be more aware of their energy consumption (Owens and Driffill, 2008; Devine-Wright and Devine-Wright, 2004).

For utilities, Strbac (2008) highlights that DSM prevents or defers the need to invest in new generation capacity, which is a costly option if the generators are used infrequently during times of peak load (and are usually carbon-intensive plants). Historically, this has been the primary reason for utility engagement with DSM outside of undertaking activities due to mandatory regulation. Torriti *et al*. (2010) extend this point by stating that due to the political pressure on the energy industry to invest in low carbon generation in a growing number of countries (such as countries in Europe, North America, and east-Asia), energy utilities have an effective way to deal with more variable sources, such as wind power, by using demand response tariffs like time-of-use pricing (described in sub-section 2.3.2). Furthermore, Strbac (2008) argues that DSM improves transmission and distribution grid investment and operation efficiency. The paper thus highlights that DSM has benefits across the total energy system.

For consumers, DSM can allow them to more actively engage in energy markets and monitor their specific patterns of energy consumption (Darby 2006, p. 3). DSM can be used as an educational tool, which may lead to cost savings through behaviour change, demand response tariffs and energy efficiency, and even the production of capital through feed-in tariffs (payments from energy utilities to consumers for each unit of low carbon energy produced) from micro-generation. Both Darby (2006, p. 3) and Stromback *et al*. (2011, p. 12) are in agreement that in the face of energy price rises in the 2010s, making energy more ‘visible’ to consumers could help them to mitigate such rises. As Capgemini (2008) notes, the success of DSM will depend on being simple, affordable, empowering and reliable for consumers.

For the market, Hirst (2002) argues that permitting and encouraging retail customers to respond to dynamic prices will improve economic efficiency, discipline wholesale market power, improve reliability and reduce the need to build new generation and transmission facilities. DSM can also facilitate the development of a ‘negawatts’ market, where reduced energy has an equal value to megawatts (Steinberger *et al*., 2009). Amory Lovins, who first coined the term ‘negawatts’, states that it refers to electricity saved and he argues for the development of a negawatts market where negawatts are treated as a commodity (Lovins, 1990). Such a market would involve a certain number of megawatts being reduced by participating customers when called on to do so by aggregators, utilities or system operators, in order to help balance supply and demand. In the industrial sector, many large consumers are involved in demand response programmes, such as interruptible/curtailment programmes, particularly in the USA, where upfront payment incentives or rate discounts are given to reduce load to pre-defined values (Albadi and El-Saadany, 2008).

It is clear that there is limited empirical evidence in the academic literature to show the transferability of the negawatts market concept to the domestic sector, where aggregators could aggregate load reductions from a number of different smaller consumers. Compared with industrial programmes, this would be a more complex administrative challenge, as there are more consumers and decision-makers to deal with. However, the market for aggregators is beginning to develop in some countries as the necessary regulatory frameworks are put in place to allow demand-side participation in balancing, reserve or capacity markets. For example, demand response can participate in a number of regional-scale capacity markets in the USA, such as PJM, NYISO, ISO-NE, MISO and CAISO. This is discussed further in chapter four.

Globally, different types of demand response programmes have been well-tested with large consumers but not to the same degree with small consumers (small commercial and residential sectors). Newborough and Augood (1999) provide the example of automation, which refers to the direct control of certain appliances and equipment by energy utilities during peak periods. They show that automation can be adapted to smaller consumers to achieve up to 60% peak-power reductions in appliance use through the controlled modulations of appliances, such as cooking and washing machines. For example, automation devices may be installed on air conditioners to change the temperature by a few degrees for a few hours during peak periods (Albadi and El-Saadany, 2007; Walawalkar *et al*., 2010). The contracts are designed so that direct load control has minimal impact on the welfare of consumers. Spees and Lave (2007) showed that 65% of firms included in their study reported that direct load control had minimal impact on facility operation. Suitable domestic equipment for automation include: fridges, washing machines, tumble dryers, lighting, air conditioning, hot water storage heaters and electric heaters.

A further example of residential demand response is Électricité de France (EDF)'s Tempo tariff, which was introduced in the 1960s. The tariff involves the days of the year being divided into three categories: 300 days are 'blue' (with low electricity prices), 43 days are 'white' (with medium electricity prices) and 22 days are 'red' (with high electricity prices) (IEA DSM, 2008). EDF informs consumers in advance of what colour the days will be based on forecasts of electricity demand, usually determined by the weather (IEA DSM, 2008) or if the French transmission network operator informs them of significant congestion on the electricity network (Crossley, 2011). However, tariff structures like Tempo are currently not very common and <20% of residential consumers in France use it (IEA DSM, 2008).

Simpler tariffs, such as higher electricity prices for peak times of the day and lower electricity prices for off-peak times, are available in some countries. Off-peak times are generally during the night, such as 10pm-6am under EDF's tariff (IEA DSM, 2008). In the UK, heat stored in bricks and electric hot water storage tanks allow residential consumers to store thermal energy during the night when electricity prices are cheaper, and use the energy during the day using Economy 7 tariffs (Hamidi *et al*., 2009) or Economy 10 tariffs, where there are seven and ten hours of off-peak electricity prices respectively. Owen *et al*. (2012) estimate that there are ~2 million households with electric storage heating and 3-3.5 million households with Economy 7 tariffs in the UK. These examples show that demand response has been useful in helping utilities to match supply and demand, but it is clear that they do not showcase the full potential of demand response.

The potential benefits of DSM are summarised in figure 3. The diagram is broken down into four boxes: benefits to the energy system, benefits to consumers, benefits to utilities and environmental benefits.

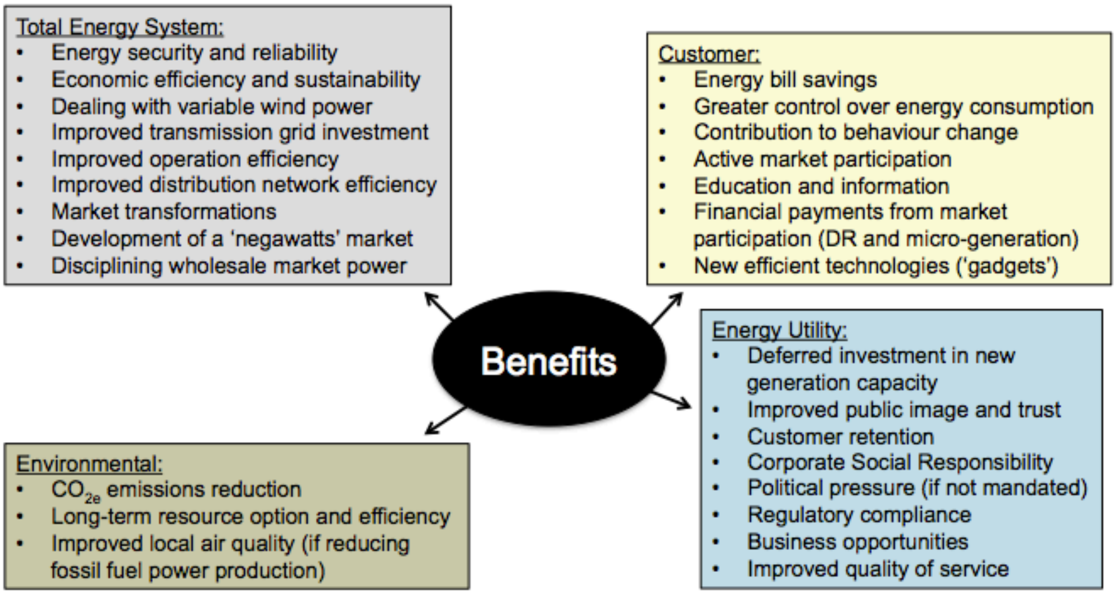


Figure 3: the benefits of demand-side management

2.1.4 Challenges

Despite the clear benefits of DSM, which are well-documented in the literature, there are a number of key implementation challenges that exist. Kim and Shcherbakova (2011) argue that the challenges are primarily social, political and economic rather than technical. From an economic perspective, Didden and D'haeseleer (2003) emphasise the implementation and incentives issues that have developed as a result of energy market liberalisation. Energy market liberalisation refers to the privatisation of the energy industry and/or the introduction of competition. Full market liberalisation involves moving away from vertically-integrated monopolies to retail competition with full customer choice of energy supplier (Bhattacharyya, 2011; Lise and Kruseman, 2008). Didden and D'haeseleer (2003) claim that *Integrated Resource Planning* (IRP) is theoretically appropriate and partly applicable in the open market, but as its history indicates, there are better frameworks for developing DSM, such as energy services and mandated energy efficiency goals. IRP is a regulatory framework that ensures that energy utilities evaluate all options (including the demand-side) for meeting future energy demands and provide energy services at minimal societal costs to customers (Cheng, 2005, p. 43). The most fundamental economic barrier to DSM development is the limited incentive for energy utilities to invest in DSM in a market based on the quantity of electricity sold if there are no clear financial returns (Cheng, 2005, p. 56), or if DSM is not cost-competitive with traditional approaches (Strbac, 2008). This barrier is explored in more depth in sub-section 2.3.3.

However, there has been growing interest in the concept of decoupling profits from the quantity of energy sold through the development of energy services and 'negawatts' markets (Steinberger *et al*., 2009). Policies and business models to achieve decoupling are discussed in sub-section 2.3.3. Consumers do not necessarily value the energy itself but the services that it can provide, such as lighting, heating, cooling, washing, cooking and mobility (Haas *et al*., 2008). Energy utilities that set up new lines of business to install measures that can achieve these services at reduced cost to the consumer, such as through energy efficiency, smart technologies and micro-generation, could profit from DSM. As Strbac (2008) argues, a concentration on energy services could help energy utilities to overcome a lack of understanding of the benefits of DSM solutions.

Nevertheless, the concept of energy services is not new and some countries, such as the USA, have being trying to develop energy services markets (particularly at a state-level) since the energy crises in the 1970s without large-scale success. Some US states, such as Vermont, have introduced separate utilities to offer energy services. In 2000, the Vermont Public Service Board appointed two entities (Efficiency Vermont and Burlington Electric Department) to provide energy efficiency services, such as technical assistance, rebates and other financial incentives (Vermont Public Service Department, 2014). It was the first energy efficiency utility of its kind in the world and other US states have now begun to copy the business model (Hamilton *et al*., 2002). Hence, there could be a role for governments to introduce the required regulatory frameworks to stimulate the market development of Energy Service Companies (ESCOs).

Similar attention is needed for the market development of demand response aggregators, whether the aggregation services are provided by separate companies or as new lines of business for current energy utilities.

Although Kim and Shcherbakova (2011) argue that the main barriers are social, political and economic rather than technical, they exclude information-based infrastructure from their category of technical issues and instead highlight it as an important limiting factor. This is strongly evident in the literature. For example, Strbac (2008) conveys how a lack of information and communications infrastructure is an important hindrance to the successful implementation of many DSM measures, such as demand response tariffs and the effective management of micro-generation technologies. In the latter case, the micro-management of a collective number of local micro-generation installations to form a 'Virtual Power Plant' (El Bakari *et al*., 2009) is dependent on the necessary information and communication infrastructure being in place. Currently, the rollout of smart meters in many countries aims to overcome this challenge in the domestic and small-commercial sectors. For example, the European Union (EU)'s *Directive 2009/72/EC* mandates that member states must achieve at least an 80% rollout of smart meters to small consumers by 2020. Smart meters are advanced energy meters that measure consumption in real-time, providing detailed information to utility companies and allowing bidirectional communication, which enables the collection of information about electricity fed back into the grid from customers' premises (Depuru *et al*., 2011), such as from solar photovoltaic panels.

Kim and Shcherbakova (2011) convey how smart meters with In-Home Displays (IHDs) (display monitors showing energy consumption, prices and other information such as carbon emissions) could overcome some of the commonly cited barriers to DSM development: consumer knowledge, consumer engagement, information feeds and two-way communication between suppliers and customers. They state that consumer knowledge refers to consumers' limited practical knowledge of the functioning of energy markets. They highlight that consumer engagement is linked to a lack of effective utility strategies to incentivise consumers to take part in demand response programmes. The paper emphasises that information feeds relate to the lowering of search costs for consumers to find information about energy prices and consumption, and what they need to do to reduce consumption if they are signed up to a demand response programme. The authors argue that IHDs have an important role to play here. Two-way communication between suppliers and customers refers to the installation of smart meters to integrate information on consumption to and from the supplier, including the use of demand response programmes and micro-generation technologies if present. Key challenges of two-way communication are the management of large volumes of data by energy utilities, as well as cyber security issues (Depuru *et al*., 2011; Chinnow *et al*., 2011). In the UK, the government introduced a Data and Communications Company to ensure the security of communications and data privacy during its rollout of smart meters between 2016-2020.

The provision and appropriateness of engaging information is arguably the most fundamental social barrier, as there is much research from behavioural economics, which shows that simply providing consumers with information (the ‘information deficit model’) does not result in the uptake of DSM measures by consumers. For example, Ürge-Vorsatz and Hauff (2001) found that, despite Hungary’s success with the market transformation of Compact Fluorescent Lighting (CFL), a high level of consumer awareness does not necessarily guarantee market success. From a theoretical point of view, Dawnay and Shah (2011) argue that people are “bad at computation” and are instead driven by other people’s behaviour, habits, doing the “right thing”, self-expectations of behaviour, being loss-averse (i.e. would prefer to avoid losing something than gaining something) and needing to feel involved to make change. In addition to these factors, family, friends and peers (‘strong ties’) are important influences on people’s decision-making (Johnson-Brown and Reingen, 1987; Granovetter, 1973), as well as the influence of neighbours or competing companies and social norms (Mau *et al*., 2008; Strachan and Warren, 2011).

The focus of smart meters in the literature has primarily been on the benefits to consumers rather than the challenges for energy utilities. Strbac (2008) argues that DSM increases the complexity of the electricity system operation when compared with traditional approaches, though this is difficult to determine until the total electricity system infrastructure for the smart(er) grid (discussed in sub-section 2.2.3) is established. It is possible that aspects of the smart(er) grid could reduce complexity instead. Nevertheless, some authors argue that DSM is too complex for utilities to take the main responsibility for management and implementation (for example, Cheng, 2005, p. 56).

The potential barriers to the development of DSM are summarised in figure 4. The diagram is broken down using the same format as figure 3 for DSM benefits: energy system, consumers, utilities and environmental barriers.

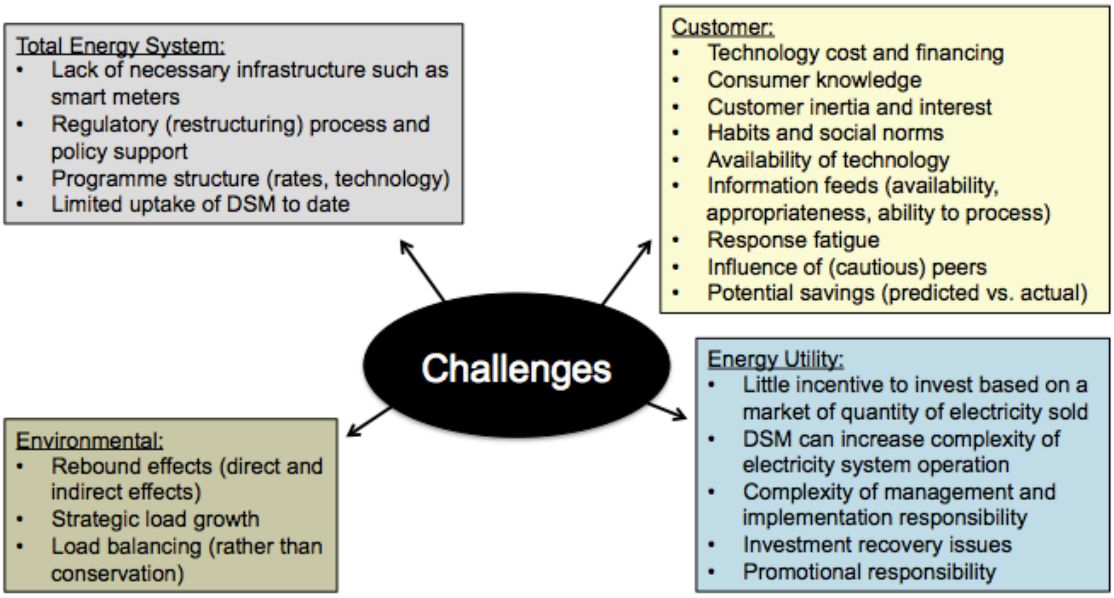


Figure 4: the challenges for demand-side management

2.2 DSM Policy History

2.2.1 DSM in National Policy

Although DSM is receiving growing research and political attention as a result of the low carbon agenda, energy security issues and the development of smart(er) grids (discussed in sub-section 2.2.3), harnessing demand-side flexibility is not new (Cooke, 2011, p. 11). The concept of DSM in national government policy can be traced back to the USA's *Energy Policy and Conservation Act* (EPCA) in 1975, which was closely followed by the *National*

Energy Conservation Policy Act (NECPA) and the *Public Utility Regulatory Policy Act* (PURPA), which were introduced as part of the *National Energy Act 1978* (McNerney, 1998, p. 27). The policies were the first instances of DSM being legislated nationally as a solution to the energy security issues of the 1970s. Nevertheless, the *notion* of DSM has been around for a long time, traditionally referring to a utility's general load management or through the use of hot water tanks and off-peak storage heaters in houses (Barrett, 2006). The latter was particularly the case in New Zealand and Europe in the 1960s and 1970s (Gellings, 1985). However, EPCA, NECPA and PURPA were the first instances of DSM in government policy.

EPCA aimed to increase energy production and supply, reduce energy demand and improve energy efficiency. For energy efficiency, the policy focused on introducing corporate average fuel economy standards for vehicles and minimum energy performance standards for equipment and appliances (Doris *et al*., 2009). However, these standards were voluntary. NECPA updated the minimum energy performance standards set by EPCA for equipment and appliances and made them mandatory (Doris *et al*., 2009). The policy also required federal agencies to perform energy surveys to reduce consumption in buildings, vehicles, equipment and operation, and provided subsidies for residential solar panels. PURPA aimed to reduce energy demand, promote a greater use of domestic energy and increase the supply of renewable energy. The policy introduced IRP, as defined in sub-section 2.1.4. Options within an IRP framework include DSM (particularly energy efficiency and demand response) in addition to traditional supply-side options and the utilities choose the least-cost combination of resources (Thomas *et al*., 2000; Cheng, 2005, p. 43). The policy also introduced rate design modifications to promote energy efficiency investments. This allowed utility incentives to be aligned with the delivery of energy efficiency (discussed further in sub-section 2.3.3). Amendments to PURPA have been made over time, notably in the *National Appliance Energy Conservation Act of 1988* (NAECA), the *Energy Policy Act of 1992* (EPAct 1992), the *Energy Policy Act of 2005* (EPAct 2005) and the *Energy Independence and Security Act of 2007* (EISA). A timeline of key energy acts and policies related to DSM in the USA, the EU and the UK since the energy crises of the 1970s are included in figure 5.

The energy crises of the 1970s particularly affected the USA and were caused by the Arab Oil Embargo of OAPEC (Organisation of Arab Petroleum Exporting Countries) in 1973-1974 and the Iranian Revolution in 1978-1979 (Hamilton, 2011, pp. 14-16). As a result, DSM programmes in the USA grew rapidly from the mid-1970s to the mid-1990s and by 1995, 600 energy utilities had conducted 2,300 programmes involving 20 million participants (Gellings, 1996; Cheng, 2005, p. 54). Notably, between 1989 and 1995, 260,000 gigawatt-hours (GWh) were saved from a cumulative spending of USD 14 billion (Nadel and Geller, 1996; Cheng, 2005, p. 54). Post-1995 DSM programmes declined in the USA as energy security issues became less prominent and Gellings (2000) argued that their future was in doubt. However, Gellings (2000)'s view has proved to be incorrect, as DSM programmes have steadily increased in the USA in the 2000s and 2010s as shown in EPAct 2005 (which covered a wide range of areas in energy policy, including energy efficiency), EISA (2007) (which introduced new performance standards for vehicles and equipment), the *American Recovery and Reinvestment Act of 2009* (ARRA) (which was introduced as a response to the *Great Recession* of 2008 and provided $17 billion in funding for energy efficiency programmes and renewable energy), and the *Energy Efficiency Improvement Act of 2015* (EEIA) (which provides the frameworks to improve energy efficiency in rented buildings, applies standards to grid-enabled water heaters, improves the energy efficiency of government buildings and improves the energy information of commercial buildings – at the time of writing this bill has not yet been enacted and is currently being considered by the Senate after passing the House of Representatives) (Yacobucci, 2014; Congress.gov, 2015).

In other continents, DSM did not achieve the same degree of development as in the USA. For example, in Europe, there was little concentration on government-stimulated DSM policy in the 1970s as energy security issues were not as prominent as those in the USA (though there were still utility-stimulated load management activities). This led to a lack of a European equivalent to EPCA, NECPA, or PURPA, which also pre-dated the formation of the European Union (EU) in 1993 (as a result of the *Maastricht Treaty* in 1992). Nevertheless, Gellings (2000) speculated that Europe had a similar degree of development as the USA from the 1980s to mid-1990s. Although energy conservation and energy efficiency measures were given more attention in political circles during this period, market liberalisation and deregulation dominated the agenda (Sorrell, 2015) and from the late-1990s, many European energy utilities' interest in DSM was removed (Wilkler, 2000; Nadel and Kushler, 2000; Waide and Buchner, 2008). Market liberalisation is defined as the transformation from monopolistic publicly owned production and distribution to privatised markets, with various competing firms (Lise and Kruseman, 2008). Although prior to liberalisation the market was based on the quantity of electricity sold, liberalisation enhanced this notion by introducing competition and thus, many energy utilities perceived energy conservation as at odds to the profitability of their businesses (Cheng, 2005, p. 56). The liberalisation of energy markets in the EU began with the introduction of *Directive 96/92/EC* (common rules of the internal electricity market) and *Directive 98/30/EC* (common rules of the internal gas market) (Kayikci, 2011). Despite energy market liberalisation, there was renewed interest in the 2000s and 2010s as a result of the growing prominence of the low carbon agenda in Europe and the role of DSM in contributing to meeting policy objectives.

Unlike other countries in Europe, the UK has followed a steady development of DSM since the energy crises of the 1970s. A number of these developments include DSM for both electricity and gas. In the 1970s, there were no national acts or policies for DSM, but government actions, such as the setting up of relevant institutions (e.g. the Department of Energy was created in 1974, the Chartered Institute of Building Service Engineers was established in 1976 and the Advisory Council on Energy Conservation was set up in 1977), and the introduction of DSM information campaigns (such as the Energy Survey Scheme (1976-1989), the Industrial Energy Audit Scheme (1976) and the Energy Conservation Demonstration Project Scheme (1978-1989)) (Mallaburn and Eyre, 2013). In the 1980s, DSM information campaigns continued and contributed to the development and passing of the Environment White Paper in 1989 ("This Common Inheritance"). In the 1990s, the influence of the EU began with the introduction of appliance energy labelling (*Directive 92/75/EEC*) and minimum energy performance standards (*Directive 92/42/EEC*) in 1992, and the EU SAVE Directive (*Directive 93/76/EEC*) in 1993, which required member states to report on energy efficiency programmes and provided funding for research and demonstration activities (Mallaburn and Eyre, 2013).

Outside of EU directives, the UK has had utility obligations on suppliers for over twenty years, and the policy model has been widely copied by other European countries (such as Italy, France, and Denmark). The first supplier obligation ran from 1994-1998 (Energy Standards of Performance (EESoP 1), the second ran from 1998-2000 (EESoP 2), the third ran from 2000-2002 (EESoP 3), the fourth ran from 2002-2005 (Energy Efficiency Commitment (EEC) 1), the fifth ran from 2005-2008 (EEC 2), the sixth ran from 2008-2012 (Carbon Emissions Reduction Target (CERT) and the Community Energy Saving Programme (CESP)), the seventh ran from 2013-2015 (Energy Company Obligation (ECO) 1), and the eighth is to run from 2015-2017 (ECO 2).

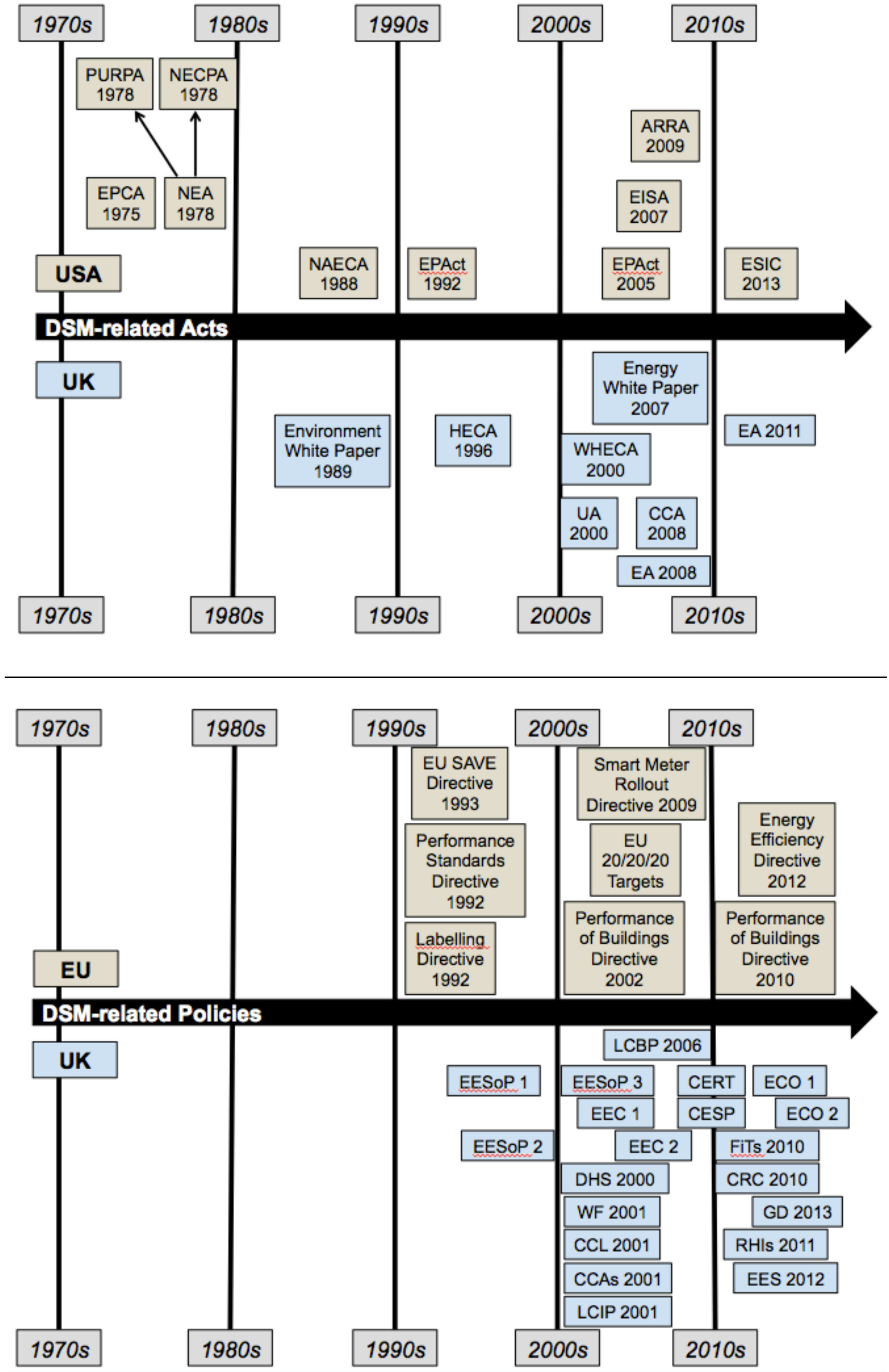
Further government actions in the 1990s include the setting up of the Energy Saving Trust in 1992 (to act as the main source of information on energy efficiency and conservation for consumers), tightening building regulations in 1995 and introducing the *Home Energy Conservation Act of 1996* (Mallaburn and Eyre, 2013). The *Home Energy Conservation Act of 1996* was the first major act dedicated to DSM and was followed by the implementation of a number of related acts in the 2000s: the *Decent Homes Standard* (2000), the *Utilities Act* (2000), the *Warm Homes and Energy Conservation Act* (2000) and the *Climate Change Act* (2008).

In the 2000s, there were a number of further important government activities and white papers, such as the Fuel Poverty Strategy (2001), *Warm Front* (2001), the Climate Change Levy (CCL) (2001), Climate Change Agreements (CCAs) (2001), setting up the Carbon Trust (2001), the *Low Carbon Innovation Programme* (2001), the *Low Carbon Building Programme* (2006), the Energy White Paper of 2007 (“Our Energy Future – creating a low carbon economy”), the implementation of the EU *Energy Performance of Buildings Directive* in 2002 (*Directive 2002/91/EC*, which introduced energy performance certificates (EPCs) and display energy certificates (DECs) for buildings), adapting building regulations in 2005 to make condensing boilers mandatory, incorporating the EU 20/20/20 climate targets (which included a 20% improvement in energy efficiency by 2020 for each EU country), setting up the Committee on Climate Change (CCC) in 2008 (to provide independent advice to the government on climate change and to set carbon budgets), setting up the Department of Energy and Climate Change (DECC) in 2008, and implementing the EU *Smart Meter Rollout Directive* in 2009 (*Directive 2009/72/EC*) (Mallaburn and Eyre, 2013).

In the first half of the 2010s, government activities on DSM continued with the introduction of the *Carbon Reduction Commitment Energy Efficiency Scheme* (CRC) in 2010, tightening the EU *Energy Performance of Buildings Directive* (*Directive 2010/31/EU*) in 2010, introducing feed-in tariffs for micro-generation in 2010 through the *Energy Act 2008*, implementing the *Energy Act 2011*, introducing feed-in tariffs for micro-renewable heat in 2011 through the *Renewable Heat Incentive*, regulations to ban F and G EPC ratings for rented commercial or domestic properties from 2018 (2011 – planned, 2015 – regulated), introducing the *Energy Bill* (2011), implementing the Energy Efficiency Strategy in 2012 (updated annually), setting up the Energy Efficiency Deployment Office (EEDO) with the responsibility of coordinating government energy efficiency policy, implementing the EU *Energy Efficiency Directive* (*Directive 2012/27/EU*), and introducing the *Green Deal* in 2013 (Mallaburn and Eyre, 2013).

It is likely that the steady development of DSM policy in the UK will continue in the second half of the 2010s. In summary, there has been a reasonably strong development of DSM policy in the UK since the 1970s and prior to the establishment of the EU, though development has been more pronounced since the 1990s and particularly since the 2000s. However, it is clear that since the 1990s the EU has been an important influence in developing national DSM policy across European countries. A timeline of key energy acts and policies related to DSM in the EU, the UK and the USA since the energy crises of the 1970s are included in figure 5. The top timeline shows the main acts introduced in the USA and the UK, and the bottom timeline shows the main policies introduced in the EU and the UK. A tabulated key is included.

Figure 5: timeline of key DSM Acts and policies in the USA, UK and EU



USA Acts	Explanation
EPCA 1975	Energy Policy Conservation Act
NEA 1978	National Energy Act
NECPA 1978	National Energy Conservation Policy Act
PURPA 1978	Public Utility Regulatory Policy Act
NAECA 1988	National Appliance Energy Conservation Act
EPAct 1992	Energy Policy Act of 1992
EPAct 2005	Energy Policy Act of 2005
EISA 2007	Energy Independence and Security Act
ARRA 2009	American Recovery and Reinvestment Act
EEIA 2015	Energy Efficiency Improvement Act (under consideration)

UK Acts and Policies	Explanation
Environment White Paper 1989	Environment White Paper
HECA 1996	Home Energy Conservation Act
Energy White Paper 2007	Energy White Paper
WHECA 2000	Warm Homes and Energy Conservation Act
UA 2000	Utilities Act
CCA 2008	Climate Change Act
EA 2008	Energy Act of 2008
EA 2011	Energy Act of 2011
EESoP 1	Energy Efficiency Standards of Performance 1
EESoP 2	Energy Efficiency Standards of Performance 2
EESoP 3	Energy Efficiency Standards of Performance 3
EEC 1	Energy Efficiency Commitment 1
EEC 2	Energy Efficiency Commitment 2
CERT	Carbon Emissions Reduction Target
CESP	Community Energy Savings Programme
ECO 1	Energy Company Obligation 1
ECO 2	Energy Company Obligation 2
DHS 2000	Decent Homes Standard
WF 2001	Warm Front
CCL 2001	Climate Change Levy
CCAs 2001	Climate Change Agreements
LCIP 2001	Low Carbon Innovation Programme
LCBP 2006	Low Carbon Buildings Programme
FiTs 2010	Feed-in Tariffs
CRC 2010	Carbon Reduction Commitment Energy Efficiency Scheme
RHIs 2011	Renewable Heat Incentives
EES 2012	Energy Efficiency Strategy
GD 2013	Green Deal

EU Directives	Explanation
Labelling Directive 1992	Directive 92/75/EEC
Performance Standards Directive 1992	Directive 92/42/EEC
EU SAVE Directive 1993	Directive 93/76/EEC
Performance of Buildings Directive 2002	Directive 2002/91/EC
EU 20/20/20 Targets	EU 20/20/20 climate targets
Smart Meter Rollout Directive 2009	Directive 2009/72/EC
Performance of Buildings Directive 2010	Directive 2010/31/EU
Energy Efficiency Directive 2012	Directive 2012/27/EU

2.2.2 Current International Experiences

Figure 5 shows that in the 2000s and 2010s there has been renewed interest in DSM in the USA, the UK and the EU. Chapter four presents evidence to identify what the main drivers are for this trend in terms of policy objectives. Nevertheless, in summary, the primary drivers tend to be carbon emissions reduction and energy security. The trend is in contrast to the predictions of Gellings (2000), who argued that the development of DSM in developed countries would continue to decline. However, recent figures from 2010 show that the combined annual utility expenditure across 18 states in the USA is >USD 900 million with annual incremental savings of ~2.8 million megawatt-hours (MWh) (Crossley, 2010).

Outside of the USA, the International Energy Agency's DSM Programme (IEA DSM Programme) has supported the advancement of DSM research globally since 1993 through a number of tasks and it aims to be the main source of DSM information and tools for governments and other institutions. Its growing database of country case studies from around the world (IEA DSM, 2004; IEA DSM 2005) highlights the increasing number of governments engaging with DSM as an alternative to supply-side solutions. Figure 6 shows the countries that have implemented and evaluated DSM policies to date. The map is developed from the literature review of 389 documents (academic papers, institutional reports and government documents) that are written in English and are accessible online. It is important to note that some countries that may have implemented DSM policies might have been excluded from the map where publications are not in English or are not accessible through the Internet. These constraints are discussed in chapter three. It is also important to note that the map shows countries that have implemented *and* evaluated DSM policies (not just those that have implemented DSM policies). The spatial distribution of DSM policy implementation and evaluation is analysed in chapter four.

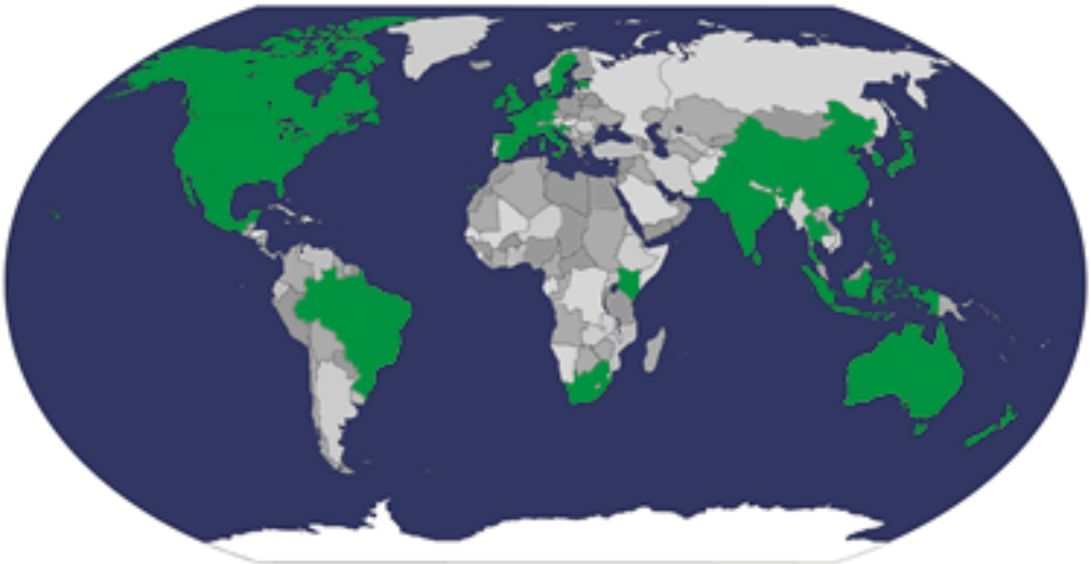


Figure 6: countries that have implemented and evaluated DSM policies

Figure 6 conveys that 35 countries and 59 states (sub-national governments) have implemented and evaluated DSM policies. The map only includes DSM policies stimulated by governments and not technological trials or DSM programmes stimulated independently by utilities. However, utility programmes that are a direct result of government policy (mandatory or voluntary) are included. An interesting temporal and spatial observation is that North American countries dominated the evaluation literature in the 1970s-1990s, European countries were prominent in the literature in the 1990s-2010s, and more recently east-Asian countries, notably China, South Korea, India and Thailand, are beginning to receive greater attention in the literature (e.g. Ming *et al*., 2013; Yu, 2012; Vine *et al*., 2006). Temporal and spatial patterns are discussed further in chapter four.

2.2.3 <u>The Smart(er) Grid</u>

There is a growing body of literature discussing the concept of the 'smart grid' or 'smarter grid'. However, as Clastres (2011) highlights, there is currently no global agreement on the definition of smart(er) grids. For example, in North America, definitions focus on energy security, resilience and reliability (Clastres, 2011), whereas in Europe, definitions concentrate on the integration of the actions of all actors in the energy system, from generators to consumers (and those that are both generators and consumers, such as consumers with on-site generation) (as discussed in Connor *et al*., 2014). Generally, smart(er) grids are discussed in terms of modernising the electricity grid by using digital and communications technology, such as smart meters and control devices (Marques *et al*., 2014). Like Connor *et al*. (2014), this research uses the definition proposed by the *European Technology Platform for Electricity Networks of the Future* (also known as the *European Smart Grids Technology Platform*):

"…electricity networks that can intelligently integrate the behaviour and actions of all users connected to it – generators, consumers and those that do both – in order to efficiently deliver sustainable, economic and secure electricity supplies". (ESGTP Strategy, 2006)

In the literature, the definition is beginning to become the standard for defining smart(er) grids, particularly in Europe, as it comprehensively captures the key aspects of the previous definitions. However, its main limitation is that it only focuses on electricity, rather than the wider energy system. Table 1 compares the traditional electricity grid with that of the smart(er) grid, which is adapted from the *European Smart Grids Technology Platform* (ESGTP) Strategy (2006).

Traditional Grid Characteristics	**Smart(er) Grid Characteristics**
Centralised control	User specified quality, security and reliability of supply for the digital age
Large generating stations	Coordinated, local energy management and full integration of DG and RES with large-scale central power generation
Technology approaching an age of one century	Flexible DSM and customer-driven value added services
Limited cross-border interconnections	Flexible, optimal and strategic grid expansion, maintenance and operation
Technically optimised for regional power adequacy	Extensive small, distributed generation connected close to end customers
Differing regulatory and commercial frameworks	Harmonised legal frameworks facilitating cross-border trading of power and grid services

Table 1: a comparison of the characteristics of traditional grids and smart(er) grids

As table 1 shows, traditional grids are primarily designed to have centralised control and large power stations, whereas smart(er) grids aim to use digital communications technology to integrate distributed generation (DG), renewable energy systems (RES) and DSM in addition to large, generating stations, and to allow two-way communication between suppliers and consumers (such as through smart meters). The key question is whether or not the concept of smart(er) grids is more effective and efficient at balancing supply and demand than traditional, centralised control.

DSM is thus one aspect of the smart(er) grid, which focuses on the actions undertaken by consumers on the demand-side of energy meters. Figure 7 summarises the role of DSM in the smart(er) grid. It uses the ESGTP Strategy (2006)'s visualisation of the smart(er) grid and adds labels to highlight examples of DSM.

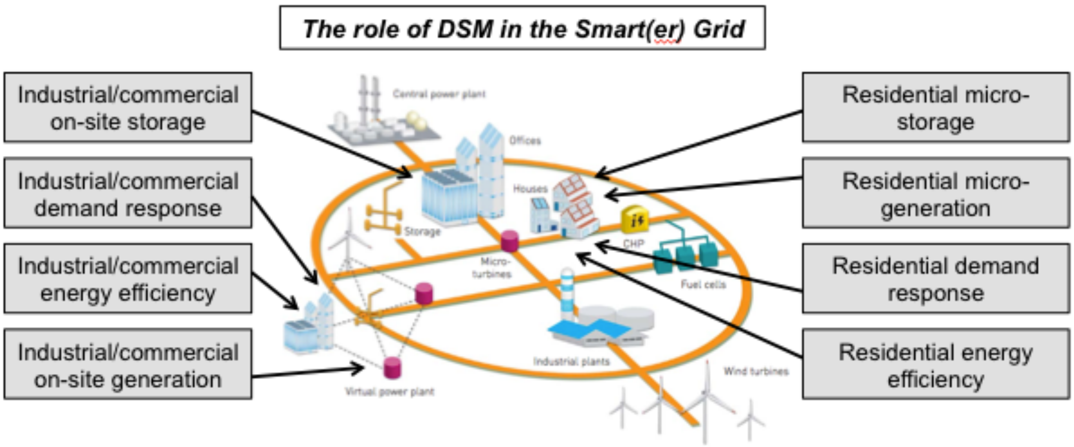


Figure 7: the role of demand-side management in the smart(er) grid

In summary, the use of on-site generation, on-site storage, energy efficiency, energy conservation and demand response in buildings (across sectors) are DSM-related activities. As discussed in sub-section 2.1.1, distributed generation, which are generation units (such as medium-sized wind farms or solar farms) plugged directly into distribution networks, is not included in the definition of DSM (except if the generation units are on-site to buildings). The same principle applies for distributed storage, which refers to storage units (such as medium-sized batteries or fuel cells) that are connected directly into distribution networks (except if the storage units, such as electric vehicles, are on-site to buildings). The micro-management of a collective number of local micro-generation and on-site generation units has been called a ‘Virtual Power Plant’ (El Bakari *et al*., 2009), which could become an important future growth area in the smart(er) grid, as new markets develop. The concept of a ‘Virtual Storage Plant’ is yet to be discussed in the literature, but it is possible for the same principles of the ‘Virtual Power Plant’ to be applied to local micro-storage and on-site storage units.

The ideas for ‘Virtual Power Plants’ and ‘Virtual Storage Plants’ revolve around a company being responsible for collectively managing a quantified amount of demand-side activities from a number of different consumers to reach a combined total (usually in MW), which can be offered into balancing, reserve, or capacity markets. This stems from the concept of the negawatts market (discussed in sub-section 2.1.3), where aggregators can collate load reductions from a number of smaller consumers and offer the combined total into balancing, reserve or capacity markets with equal value to supply-side options (Steinberger *et al*., 2009). Here, consumers are paid not to use energy for non-essential activities during times of high unit cost (and carbon), such as during peak times, when called on to do so by the aggregator, which is a type of demand response activity (Albadi and El-Saadany, 2008). Demand response markets are beginning to develop in Europe, though they have had more established success in North America, such as through regional level capacity markets in the USA, for example, PJM, NYISO, CAISO and ISO-NE (which are discussed further in chapters four and five). Despite this, the markets for ‘Virtual Power Plants’ (or ‘Virtual Storage Plants’) are underdeveloped globally.

Section 2.2 has provided a discussion of the policy side of DSM in terms of its history in national policy since the 1970s, current international experiences and its role in the future smart(er) grid. Section 2.3 delves into public policy theory before discussing DSM policy theory in more detail.

2.3 Policy Theory

2.3.1 The Policy Process

Policy research can be divided into studying the policy process, policy design, policy implementation and policy evaluation (Hill, 2009; Nagel, 2002). Policy implementation and evaluation form key parts of the research, which are discussed in chapter three. In political science, there is a vast literature on policy evaluation, and in the *Stages Model* of the policy process, outlined in Hogwood and Gunn (1984, p. 4), policy evaluation is the eighth stage in a nine-stage process, as shown below:

1. Deciding to decide
2. Deciding how to decide
3. Issue definition
4. Forecasting
5. Setting objectives and priorities
6. Options analysis
7. Policy implementation, monitoring and control
8. Evaluation and review
9. Policy maintenance, succession and termination

However, the *Stages Model* has been criticised in the literature for being overly simplistic of how policy works (for example, Hill, 2009, p. 143; John, 1998, p. 196). Hill (2009) describes evaluating policy as a function of a controlled trial method (with a control group) or reaching desired states (such as a reduction in air pollution) (pp. 279-280). These are important criticisms but they do not form the only methods of determining policy success. Undertaking randomised controlled trials (RCTs) are much needed in the energy policy field, as few have been undertaken to date compared with other disciplines, such as the medical sciences (for example, see Akobeng, 2005). However, the primary reason for this is the scale of evaluating a national policy compared with a small-scale trial. RCTs require both a treatment group and a control group, which contain participants that were randomly chosen to participate in the study from the general population and were then randomly assigned to each group (Kellstedt and Whitten, 2013, pp. 76). Furthermore, the RCTs need to be double blind with participants and evaluators blind to participants' study groups (Akobeng, 2005).

Thus, undertaking RCTs at a national policy level would need to compare the impacts of the policy with either the circumstances before the intervention (the baseline) or comparing the impact with another country that has not had the intervention. However, both cases are fraught with issues of determining cause and effect, as there are too many confounding variables that could have caused the identified relationship at this scale. In the first case, the baseline cannot be accurately established as it is almost impossible to replicate the experiment under exactly the same circumstances due to a range practical, ethical, economic and political challenges. In the second case, differences in context between countries (such as pricing regimes, political structures and cultural attitudes) reduce the reliability of the results, as each contextual factor is an important confounding variable in explaining the identified relationship. Context is discussed further in chapter six.

In contrast to Hill (2009), Nagel (2002) defines policy evaluation as: "…evaluating alternative public policies, as contrasted to describing them or explaining why they exist" (p. 133). Nagel (2002) highlights a number of criteria for evaluating government policy: equity, validity, importance, usefulness, originality and feasibility (p. 134-136). He extends this to include the following broad terms: effectiveness, efficiency, public participation, predictable and procedural due process, and political feasibility (p. 92). He describes an effective and specific method for evaluating policy called multi-criteria decision-making (MCDM) analysis, also known as multi-criteria (mapping) analysis. MCDM is a quantitative method (though can contain some qualitative aspects) and takes place in face-to-face interviews with experts. It involves the listing of policy alternatives and the judging of criteria by the participants, which are then scored to obtain a summation score that ranks the policy alternatives in a list of the participants' preferences (Nagel, 2002, p. 155). This method is utilised in the research as a secondary method and is discussed in more depth in chapter three. Where RCTs cannot be conducted due to ethical, practical and economic reasons, MCDM analysis can be an effective tool to quantitatively determine the success of policies. However, it is still based on the judgements of participants and suffers from a number of reliability issues that are not present in RCTs. The strengths and limitations of MCDM analysis are discussed in chapter three.

In addition to policy evaluation, Nagel (2002) and Hill (2009) detail the importance of other stages in the policy process, notably policy design and implementation. Governments have a number of 'tools' at their disposal for the implementation of policies (John, 2011, p. 10). Figure 8 summaries the tools, which are adapted from John (2011).

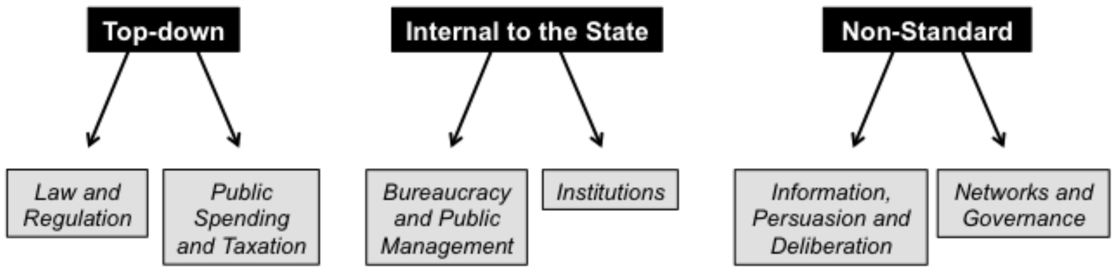


Figure 8: the 'tools' of government

Historically, much of the focus in energy policy implementation has been on the role of tools that are top-down, such as regulations, or internal to the state, such as the use of existing and new institutions (John, 2011). However, there has been a growing interest in the role that non-standard tools, such as persuasion and deliberation, can play in addition to traditional methods. John (2011) provides an excellent overview of recent discussions on the general policy process, though a criticism of his categorisations is the use of the label 'non-standard' for some tools that have been used as 'standard' policy approaches in the past, particularly in energy policy, such as information campaigns and stakeholder networks.

In the UK, more attention is starting to be given to the role of information, persuasion and deliberation in policy implementation. For example, the UK Government set up a Behavioural Insights Team within its Cabinet Office in 2010 with the aim of applying the 'nudge' theory to areas of policy, such as energy efficiency, health, tax, consumer data access and organ donation

(Behavioural Insights Team Annual Update 2010-2011 report). The UK is the first country to set up a dedicated team within Government to apply behavioural theories to public policy. However, the Behavioural Insights Team was privatised in 2014 to become independent of government. The 'nudge' theory was developed by Thaler and Sunstein (2008) and argues against the traditional neo-classical economics view that people always make decisions rationally under conditions of being fully informed through access to perfect information. Instead, it states that people make predictable mistakes based on their experiences (heuristics) and social interactions. Dawnay and Shah (2011) (in Dietz *et al*., 2011, pp. 74-75) expand this notion by listing key influences on people's decisions:

- Other people's behaviour matters
- Habits are important
- People are motivated 'to do the right thing'
- People's self-expectations influence how they behave
- People are loss-averse
- People are bad at computation
- People need to feel involved

The influence of peers, experiences, habits, social norms and availability of (and ability to process) information are key characteristics of human behaviour (Strachan and Warren, 2011) that need to be considered in the design and implementation of policies to encourage consumer engagement with DSM. As Gellings and Chamberlin (1993) note: "Research indicates energy-use behaviour and belief are resistant to change. Successful approaches to the consumer seem to be personal, possibly emotional, specific and narrow with concrete suggestions" (p. 340). Hence, the traditional incorporation of rational choice theory (the theory stating that people act rationally) in policy development is beginning to be replaced with a more holistic policy framework (notably in European countries) that incorporates the importance of information, persuasion and deliberation. These arguments have particular relevance to the residential sector, but can similarly be applied to the non-residential sector. For example, competitors may represent credible 'peers' and hence an organisation is more likely to adopt certain measures if their competitors do (Gellings and

Chamberlin, 1993, p. 334). A recent, comprehensive review by Sorrell (2015) highlights that governments are increasingly using ideas from behavioural economics and social psychology to inform policy design for energy efficiency, though argues that an effective policy approach would draw upon all disciplinary perspectives.

Hogwood and Gunn (1984)'s *Stages Model* is useful in breaking down the policy process into sub-stages beyond simply design, implementation and evaluation. However, an important criticism is that evaluation should take place at all stages of the policy process. Furthermore, the post-policy evaluation should be used in the design of future policies. The next sub-section applies these ideas to the discussion of DSM policy theory.

2.3.2 DSM Policy Types

DSM policies can be categorised into those that are market-based, regulatory, fiscal, information-based or voluntary (Warren, under review). Some authors, such as Grubb (2014) reduce the number of policy categories to three: information-based tools, regulatory standards and financial incentives (p. 165). Here, market-based policies would mainly come under financial incentives and voluntary policies would primarily come under information-based tools or financial incentives. Nevertheless, five categories are used in order to provide more depth to the analysis. However, in practice, many DSM policies cut across different categories, for example, the UK's *Green Deal* policy (implemented in 2013) has both financial and voluntary policy aspects to its design – it is voluntary for consumers to take advantage of the scheme and it has a financing mechanism that removes the upfront costs of energy efficiency purchases by recovering payments through the savings generated (UK DECC, 2010). A further important consideration is that in practice, demand-side policies can be implemented as a policy package rather than as stand-alone policies (explored in chapter four). Figure 9 summarises the main categories of DSM policy and table 2 then lists specific policies for each category.

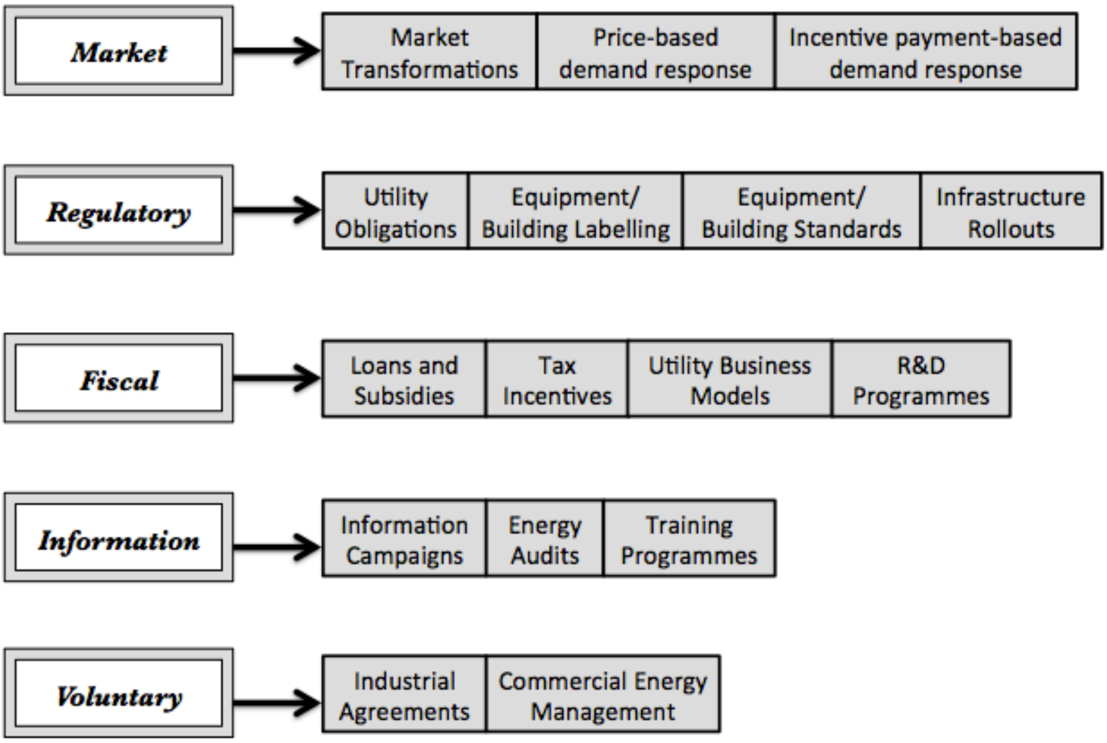


Figure 9: the main categories of demand-side management policy

Market-based DSM policies include market transformations (MT), incentive payment-based demand response (IPBDR) and price-based demand response (PBDR). Market transformations refer to long-term broad policies for stimulating the market development of energy efficiency and other demand-side resources, as well as overcoming the market barriers to DSM technological development and innovation. Although a more specific definition of MT is lacking in the energy policy literature, some authors have proposed that it should have a more technical meaning that relates to the encouragement of social, technological and economic change in the direction of greater energy efficiency (Blumstein *et al*., 2000; Geller and Nadel, 1994). From a policy perspective, MT can include a range of other DSM policy categories, such as subsidies, information campaigns, performance standards and labelling programmes that are usually included together as policy packages, but the main focus is on overcoming the barriers to DSM development, which is not the specific focus of the other DSM policy categories as stand-alone policies.

Policy Category	Specific Policy
Incentive payment-based demand response	Direct load control (automation)
	Interruptible/curtailable programmes
	Demand bidding/ancillary services market provision
	Emergency demand response
	Relieving network constraints in specific locations
Price-based demand response	Time-of-use pricing
	Critical peak pricing
	Real-time pricing
	Extreme day pricing
	Extreme day critical peak pricing
	Inverted block pricing
Market transformations	Energy efficiency barriers removal/market stimulation
Infrastructure rollouts	Smart meter rollouts (with/without In-home displays)
Utility obligations	Utility obligations
	Energy efficiency white certificate trading systems
Labelling	Appliance labelling
	Equipment labelling
	Building labelling
Performance standards	Appliance standards
	Equipment standards
	Building standards/codes
Loans and subsidies	Loans (no- or low-interest)
	Subsidies/rebates/grants
	Tax reductions/credits/exemptions
Utility business models	Decoupling (no cost-recovery)
	Shareholder incentives: shared benefits
	Shareholder incentives: performance targets
	Shareholder incentives: rates of return
	Revenue regulation
	Cost-recovery mechanisms
	Direct incentives/payments
	System benefits charges/public goods charges
Research and development programmes	Government-stimulated large-scale DSM technological trials/pilots
	Market development and deployment programmes
Information campaigns	General information and marketing programmes
	Energy auditing programmes
	Information centres
	Education, training, certification and technical support
	On-bill information
	Leading by example (e.g. public sector)
Voluntary programmes	Agreements with industrial companies
	Agreements with large commercial organisations

Table 2: the main specific demand-side management policies

Incentive payment-based demand response refers to the response of consumers to incentives (excluding price changes), such as direct financial payments for reducing non-essential energy consumption during peak times (Albadi and El-Saadany, 2008). From a policy perspective, it refers to the presence of regulatory frameworks to allow the inclusion of demand response as a resource in balancing, reserve, or capacity markets. Most IPBDR policies to date have focussed on industrial agreements for peak load reduction, rather than the aggregation of load reductions from a number of smaller loads (as discussed in sub-section 2.2.3).

Price-based demand response refers to the response of consumers to price changes, such as consuming when energy prices are cheap (e.g. during the night) and reducing consumption during peak times (e.g. in the morning when people wake up or in the evening after work when people are cooking and watching television) (O'Connell *et al*., 2014). From a policy perspective, like IPBDR, it refers to the presence of regulatory frameworks to allow the introduction of different consumer tariffs, such as time-of-use pricing (where it is cheaper to consume at certain times of the day or year), critical peak pricing (where it is expensive to use energy during peak periods) or real-time pricing (where consumers are charged hourly fluctuating prices reflecting the real cost of energy in the wholesale market) (Albadi and Al-Saadany, 2008).

Regulatory policies include utility obligations (UO), infrastructure rollouts (IR), performance standards (PS) and labelling (LB). Utility obligations refer to the setting of energy savings or carbon savings targets for suppliers, distributors, public entities, or building owners or users (THINK, 2012). The obligated parties must meet quantitative targets by specific dates by delivering or procuring eligible energy savings produced by implementing approved end-use energy efficiency measures (IEA DSM Programme, 2012). From a policy perspective, UO involves the setting of mandatory targets that are usually backed by legislation and regulation, and involve penalties for utilities that miss their targets.

Infrastructure rollouts refer to the national or regional deployment of DSM-enabling infrastructure, such as smart meters and in-home displays (monitors that display information about energy consumption, costs, carbon emissions and other relevant information). From a policy perspective, smart meter rollouts require appropriate regulatory frameworks and legislation to ensure that rules are adhered to, such as smart meter design, ensuring data privacy, detailing the timeline for the rollout, assessing costs and likely impacts, and clearly stating which parties (usually utilities) will undertake the rollout (Cuijpers and Koops, 2012; McKenna *et al*., 2012).

Performance standards refer to minimum energy performance standards for appliances, equipment and buildings. From a policy perspective, PS can be mandatory or voluntary and apply to all relevant manufacturers of appliances and equipment. Building regulations and building codes apply to all relevant building owners, managers and developers. In most cases, the standards are revised over time to reflect improving levels of energy efficiency (Harrington and Damnics, 2001). Mandatory standards are backed by legislation and regulations (Lee and Yik, 2004), and involve penalties for non-compliance. Voluntary standards set targets or negotiated agreements for average market efficiency rather than specifying the performance of individual appliances or equipment (Harrington and Damnics, 2001).

Labelling refers to the inclusion of relevant information on energy performance on appliances, equipment and buildings. Information is usually displayed within a label (for appliances and equipment) or certificate (for buildings) and includes information such as energy efficiency, carbon savings and cost savings. From a policy perspective, like PS, LB can be mandatory or voluntary and aims to encourage the consumer uptake of energy efficient technologies to contribute to meeting policy objectives for environmental protection (Banerjee and Solomon, 2003). Labels can be mandatory (backed by legislation and regulation with penalties for non-compliance), or voluntary (where manufacturers and developers are not obliged to display labels or certificates). Although labelling is an information-related policy, it is separated from the information campaigns category in order to emphasise that the latter group focuses on targeted campaigns and energy audits, rather than larger scale labelling of entire product chains and buildings. Furthermore, labelling is often implemented alongside performance standards (as shown in chapter four), which similarly requires regulation to implement.

Fiscal policies include loans and subsidies (L&S), tax incentives, funding for research and development programmes (R&D) and alternative utility business models (UBM). Loans and subsidies (including no- or low-interest loans, grants, rebates and payments) focus on providing capital for (usually specified) DSM technologies (e.g. see Gillich and Sunikka-Blank, 2013). Tax incentives (including tax reductions, tax credits and tax exemptions) concentrate on reducing the costs of DSM technologies by reducing the tax paid on purchasing them directly or indirectly (through reductions in other (non-energy-related) taxes) (e.g. see Gold and Nadel, 2011). A Bonus-malus tax system involves increasing taxes for energy inefficient technologies in addition to reducing taxes for energy efficient technologies. In this research, tax incentives are included within the broader category of ‘loans and subsidies’, as outlined in chapter three.

Research and development programmes refer to the government funding of large-scale research and innovation activities or to reduce the costs of certain DSM technologies through development, demonstration and deployment (e.g. see Kimura, 2009). Only large-scale R&D programmes are included in the research, which is justified in chapter three. Although figure 9 includes R&D programmes in the fiscal category, it can cut across different policy types (for example, it could be included under the information-based category). Here, it is included in terms of government funding.

UBM refers to alternative business models for energy utilities, such as cost-recovery mechanisms, decoupling policies, energy service company models (ESCOs) and negawatt aggregators, to put demand-side options on an equal basis (in terms of profitability) to supply-side options (e.g. see NRDC, 2013). UBM has been implemented widely in the USA at a state-level but has had limited application in other countries. Decoupling policies aim to decouple the amount of profit that utilities make from the amount of energy that they sell (Hayes *et al*., 2011), in order to overcome the disincentive for utilities to invest in energy efficiency as a resource. This can be achieved through cost-recovery mechanisms (Crossley, 2010), which are discussed in sub-section 2.3.3. Complementary approaches to cost-recovery mechanisms are the development of new business arms for utilities to offer negawatt aggregation services (discussed in sub-section 2.1.3) or energy efficiency services through ESCOs.

ESCOs aim to focus on the services that are provided by the use of energy rather than simply the amount of energy that is bought and consumed. Thus, through ESCOs, utilities can have the incentive to provide the services using less energy. Services include heating, cooling, lighting, comfort and refrigeration (Boait, 2009). Although there are a number of ESCO business models, commonly the ESCO would fund the installation of energy efficiency technologies in the consumers' buildings and the consumers would pay back the costs (plus an interest rate) through their energy bills. Performance contracting ensures that services are delivered by monitoring performance (Boait, 2009), and guaranteeing that the savings produced from DSM measures are sufficient to cover the cost of the measures over their lifetime (ICF International, 2007). UBM is explored in more depth in sub-section 2.3.3. It is included under the broader fiscal group due to the focus on financially incentivising utilities to place demand-side options on an equal basis with supply-side options. However, there is some overlap with the regulatory group, as regulation is required to implement the policy.

Information-based policies include information campaigns, energy audits, and education and training programmes. In the research, all three categories are included under the broader category of information campaigns (IC), as justified in chapter three. IC refers to government funding for general marketing campaigns conducted either directly by the government (or a government agency) or by a third party. The campaigns aim to encourage the adoption of DSM measures by consumers and/or to educate them on energy issues and the role that DSM can play (e.g. see Murray, 2010). Energy audits aim to provide information to consumers on their consumption and how to manage and reduce it in order to reduce costs. They are site-specific programmes (though they can be conducted for a number of sites where consumers have multiple buildings) and recommendations are tailored to the consumer in question (e.g. see Fleiter *et al*., 2012). Education and training programmes aim to develop the skills base for DSM by establishing formal certification systems and qualifications for energy management and auditing. Examples include the ISO 50001 international standard for energy management and the ISO 140001 international standard for environmental management (ISO, 2014).

From figure 9, twelve DSM policy categories are analysed in the research: incentive payment-based demand response (IPBDR), price-based demand response (PBDR), market transformations (MT), utility obligations (UO), performance standards (PS), labelling (LB), infrastructure rollouts (IR), loans and subsidies (L&S), alternative utility business models (UBM), research and development programmes (R&D), information campaigns (IC) and voluntary programmes (VP). Tax incentives are included under L&S, energy audits and education and training programmes are included under IC, and industrial agreements and commercial energy management are included under VP. Chapter three discusses and justifies the policy categories further.

2.3.3 Policy Challenge: Incentivising Utilities

Sub-section 2.1.4 highlighted the main challenges and barriers to the development of DSM. Two of the greatest DSM policy challenges are how to incentivise utilities and consumers to engage with the demand-side. For utilities, there is a disincentive for them in selling less of their product by investing in energy efficiency in their customer base. For consumers, engaging with demand-side activities has often been seen as an area of little interest and excitement to them, as it is not a priority issue in the context of other everyday concerns of a more social nature (Whitmarsh, 2011; Poortinga and Pidgeon, 2003). As discussed in sub-sections 2.1.4 and 2.3.1, many consumers exhibit ‘wider rationalities’ where they are influenced by the actions of peers, neighbours, family, social norms and habits (Strachan and Warren, 2011; Dawnay and Shah, 2011; Johnson-Brown and Reingen, 1987). Despite this, this sub-section focuses on how to incentivise utilities, as there is a vast literature on consumer decision-making and how to incentivise consumers (e.g. Wilson and Dowlatabadi, 2007; Granovetter, 1973; Mau *et al.*, 2008).

Utilities in most countries, particularly those with liberalised energy markets, operate in a market where profits are based on the amount of electricity or gas sold. Historically, DSM activities undertaken by utilities were for load management purposes (primarily utilising demand response to reduce costs), to comply with regulation or to defer investment in infrastructure (such as new generation capacity). There are also a number of indirect benefits to utilities from engaging with the demand-side, such as customer retention and improved corporate image and trust. However, the direct financial benefits, such as through new business opportunities (energy services and negawatt aggregation services), have been reasonably limited outside of North America (where contextual factors, such as the market structure, climate, energy demands, electricity systems, culture and regulatory environment are different to that of other regions, such as Europe and east-Asia). The primary reason for this is that current utility business models and regulatory frameworks are not aligned with allowing demand-side options to be profitable on a competitive basis with supply-side options.

The discussions in this sub-section focus on market structures that are not fully liberalised and how demand-side options can be incentivised in such market environments. The key future challenge for policy is how (if it is indeed possible) to apply these ideas to liberalised markets. One of the most important aspects in this respect is utility structure. Vertically integrated utilities own the supply, transmission and distribution aspects of one energy resource (electricity, gas or water), whereas horizontal utilities own more than one energy resource (electricity, gas or water) but for one aspect of the energy system (supply, transmission or distribution) (Joskow, 2008; Walsh and Todeva, *n.d.*). However, in practice, a number of liberalised markets have a mixture of both types of utility. For example, in the UK, the six largest energy companies (British Gas, EDF Energy, E.ON UK, npower, Scottish Power, and Scottish and Southern Energy) have both generation and supply arms, and are involved in both the electricity and gas markets. In practice, the companies generate power and sell it to their supply arms rather than trading it on the open market (Ofgem, 2014).

Transmission and distribution networks are natural monopolies and National Grid operates the transmission networks in the UK. There are eight main distribution network companies in the UK, which operate on a geographical basis. Unusually, two of the six largest energy suppliers (Scottish Power and Scottish and South Energy) are also distributors. Natural monopolies exist in many states in the USA and it is here that experiences with alternative utility business models (UBM) have had practical success. The country has been the pioneer in introducing policy and business models to incentivise utilities to engage with the demand-side. However, natural monopoly utilities are subject to economic regulation where regulatory state authorities, such as public utilities commissions, set the 'rates' (the prices that consumers are charged). In the USA, public utilities are either publicly traded corporations or private businesses. In liberalised energy markets, such forms of economic regulation no longer apply and prices are usually set through competition (RAP, 2011). Thus, the challenge for implementing UBM in liberalised markets becomes apparent and this is the subject of current work that is forthcoming (in Warren, under review).

The business and policy models for UBM in markets with public utilities are summarised in table 3, which is adapted from Hayes *et al*. (2011), Crossley (2010) and Warren (under review). Business models focus on the utility and policy models focus on the required regulatory support.

Business model	Decoupling (without recovery)
	Lost Revenue Adjustment Mechanisms (LRAMs)
	Shareholder Incentives: Shared Benefits
	Shareholder Incentives: Performance Targets
	Shareholder Incentives: Rates of Return
Policy model	Revenue regulation
	Full recovery (reduced sales and DSM programme costs)
	Enhanced full recovery (plus additional profit)
	Recovery of reduced sales only (partial recovery)
	Recovery of DSM programme costs only (partial recovery)
	Direct incentives for DSM

Table 3: business and policy models to incentivise utilities to engage with DSM

In terms of business models, decoupling seeks to detach the amount of profit a utility makes from the amount of electricity or gas it sells by ensuring that it can benefit from offering demand-side services to its customers. This is done through the regulation of its revenue by making sure that it does not receive less than the authorised fixed costs nor more from increased sales (Hayes *et al*., 2011). Thus, the disincentive to engage with DSM is removed and the incentive to sell more electricity or gas is removed. Lost Revenue Adjustment Mechanisms (LRAMs) allow utilities to recover lost revenue through rate adjustments, though they can still increase returns from increased sales (Baxter, 1995). Thus, this is a form of partial decoupling, as it deals only with revenue losses (Eto, 1994). Shareholder incentives aim to take the previous concepts a step further by allowing utilities to profit from DSM programmes and they can be split into three main types: shared benefits, performance targets and rates of return (Hayes *et al*., 2011). Shared benefits involve the utility sharing the benefits of DSM programmes with the ratepayers, with the utility benefiting from some of the value of the energy savings achieved. Performance targets aim to reward utilities for achieving previously defined energy or carbon savings goals through energy efficiency. These targets are often set on a 1-3-year basis. There is overlap here with utility obligations (UO) (called energy efficiency resource standards in the USA), as targets are set for utilities. The main difference is that these mechanisms try to adapt or change the underlying business models of utilities rather than simply set targets that utilities must meet. Furthermore, rewards are not given for successfully meeting targets (though penalties usually exist for not meeting targets). Thus, UO does not aim to change the business models of utilities, which is where the definitional boundary is drawn. Rates of return seek to give utilities a return based on efficiency spending or savings (Hayes *et al*., 2011).

In terms of policy models, revenue regulation reflects the decoupling business model to ensure that utilities receive consistent revenues regardless of the amount of electricity or gas that they sell but with periodic (often annual) adjustments to the revenues based on actual sales (to prevent under- or over-selling in any given year) (Crossley *et al*., 2000). The recovery of foregone revenue (from reduced sales) and DSM programme costs policy model reflects those business models that involve cost-recovery mechanisms. With full recovery, the utility can recover reduced sales and DSM programme costs (Crossley, 2010). With partial recovery, the utility can recover either reduced sales or DSM programme costs (Hayes *et al*., 2011). With enhanced full recovery, the utility can recover reduced sales, DSM programme costs and additional profit (Warren, under review). Direct incentives for DSM is a policy model that reflects the shared incentives business models, but also includes direct payments to utilities for DSM activities (Crossley, 2010)

Figure 10 graphically summarises the main business and policy models for incentivising utilities to engage with the demand-side and how they impact on utility revenues and costs. Utility costs include investment in new generation capacity, infrastructure, operation and maintenance. In each business and policy model (1-4), utility costs decrease overall (theoretically by the same amount), but this is primarily due to deferred investment in generation infrastructure and improved efficiency of operation rather than general operation and maintenance (as DSM can sometimes increase utility costs due to increased administrative complexity). The *x*-axis shows the independent variable (UBM) and the *y*-axis shows the dependent variable (utility revenues and costs). The general premise is that moving from 1-4 along the *x*-axis results in a greater difference between revenues and costs as utility revenues increase. In other words, the benefits to the utility increase as one moves from 1-4. However, the funding source for the additional benefit needs to be established, which is usually provided through consumer bills though governments may directly pick up the costs if the impact on consumer bills is too great.

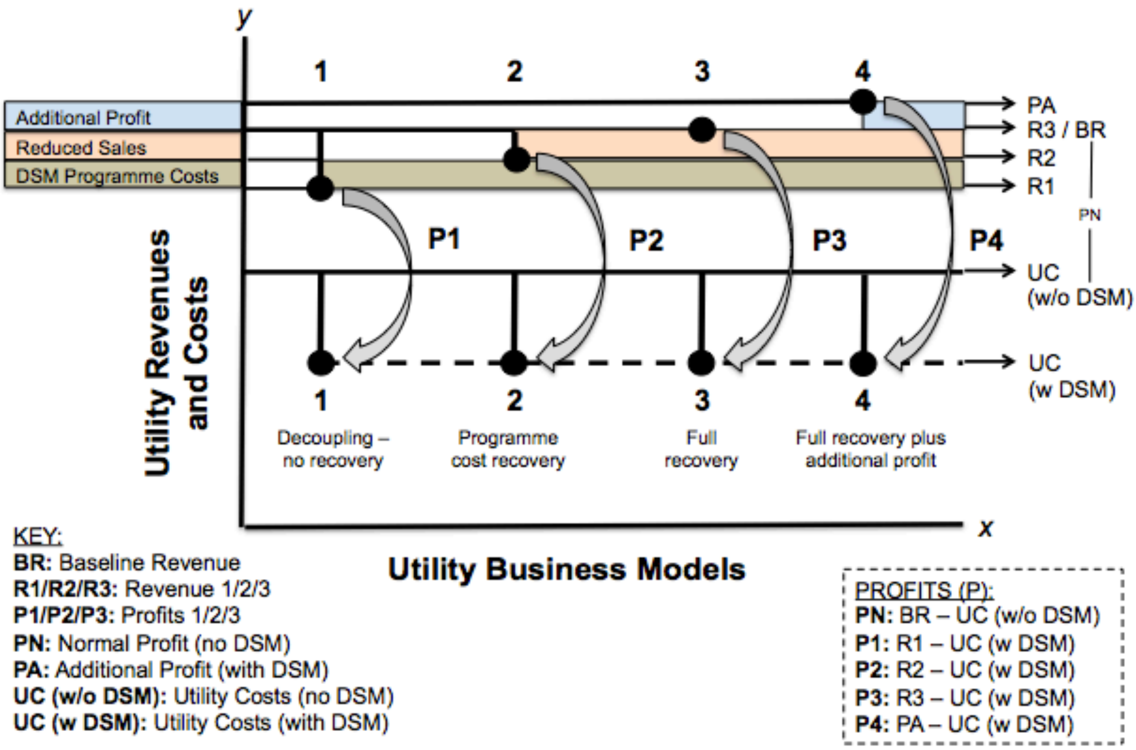


Figure 10: business and policy models to incentivise utilities to engage with DSM

As figure 10 shows, introducing decoupling policies without any recovery of potential lost revenue from reduced sales and DSM programme costs (UBM 1) leads to a reduction in utility revenue but still leads to an overall benefit as the reduced utility costs are greater than the reduced sales. In this case the utility funds any additional costs. Introducing policies to recover DSM programme costs (UBM 2) increases the utility benefit as it can recover the administrative and deployment costs of DSM programmes. This is partial cost recovery – LRAMs fit into this category in reverse where potential lost revenue from reduced sales can be recovered but not DSM programme costs. Full recovery allows the recovery of DSM programme costs *and* potential lost revenue from reduced sales (UBM 3). Enhanced full recovery (UBM 4) allows the full recovery of costs as in (3) but allows extra profit to be recovered in addition. Some types of shareholder incentives fit into this category. In all cases (1-4), costs are recovered primarily through consumer bills.

The practical implementation of such business and policy models depends on the individual contexts of the country or state where they are being introduced. Despite this, there has been limited practical implementation of UBM outside of the USA. Although implementation has been driven at a state-level rather than at the national-level in the USA, a growing number of states have introduced the policies within their jurisdictions. In 2013 half of the states had introduced decoupling policies, as shown in figure 11 overleaf (taken from the Natural Resources Defense Council (NRDC), 2013).

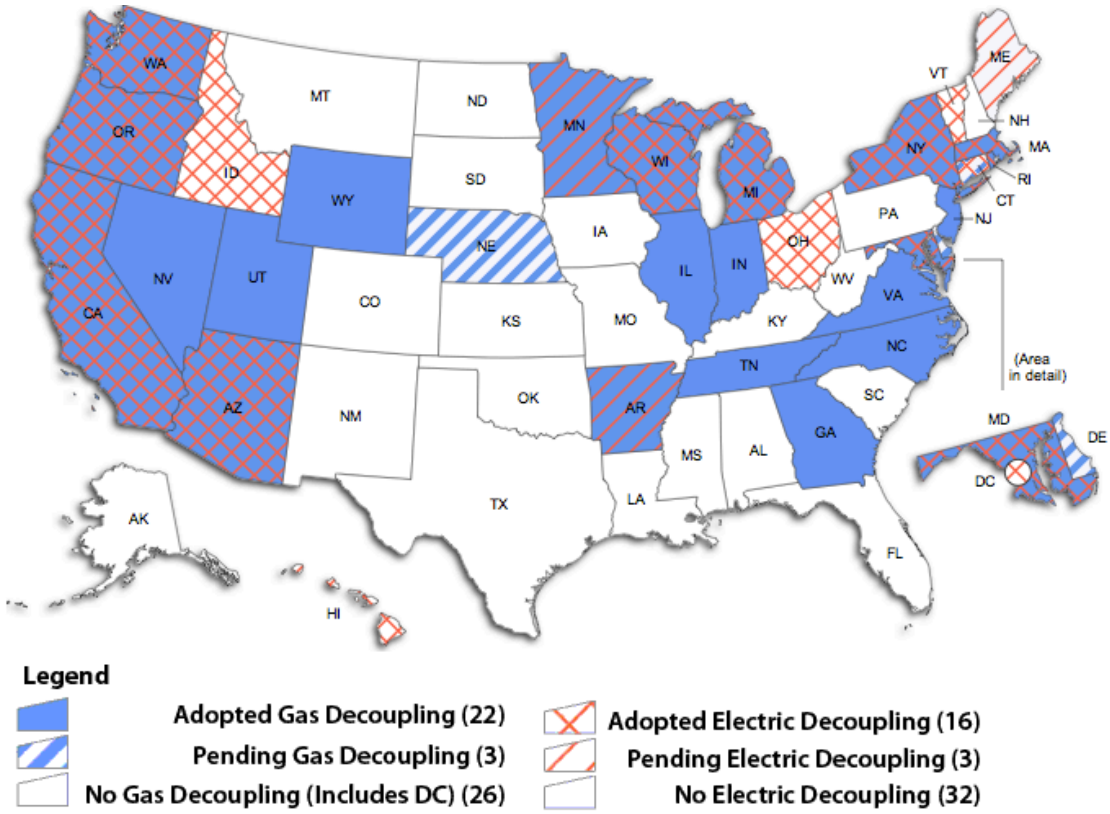


Figure 11: US states that have introduced decoupling policies for electric and/or gas utilities as of August 2013

Twenty-six states plus the District of Columbia had some form of utility decoupling (electric and/or gas) in place by August 2013 (NRDC, 2013). In addition, three states had pending gas decoupling (Nebraska, Delaware and Connecticut) and three had pending electric decoupling (Minnesota, Arkansas and Maine). Although it is unlikely that decoupling will be implemented at a national level due to differences in context between the states, it is likely that most states will individually implement some degree of decoupling during the 2010s.

The US experiences with decoupling policies have generally been successful, but a key challenge is in determining their transferability to other continents, such as countries and states in Europe, Asia, Australasia, South America or Africa. In the USA, a large number of utilities operate locally or regionally and are usually publicly- or investor-owned but with a strong degree of regulation from state public commissions. In Europe, utilities are often large, private and multi-national with operations in different countries. Thus, a policy implemented in one country of operation should be implemented in the other countries where the utility operates in order to prevent the potential misalignment of business models to the regulatory frameworks of the countries in question. The European Commission could be an effective way to deliver decoupling policies across European countries, though this would still be challenging due to the contextual differences between the USA and Europe (and within Europe), such as market structures, climate, energy demands, electricity system structures, culture and regulatory environments. Large utilities operating in countries with fully liberalised energy markets can be adverse to a strong degree of government intervention and increased regulation, which can cause political challenges for those governments. This is due to the strong focus on developing free, competitive markets in Europe (e.g. see European Commission, 2012) and UBM requires the regulation of utilities. In east-Asia, especially China, there could be more opportunity for successfully transferring decoupling policies from the USA, as utilities have a strong governmental presence.

2.4 DSM Policy Evaluation

2.4.1 Theory

Rigorous policy evaluation should include a critical appraisal of the full policy process from policy proposal through design and implementation to the post-policy evaluation stage, as argued in sub-section 2.3.1. Policy theory-based approaches emphasise the need for evaluation during the implementation stage of the policy (see Rossi *et al*., 2004; Rogers *et al*., 2000). They aim to identify issues that occurred in the design and implementation in an iterative process of design, evaluation and redesign (Harmelink *et al*., 2008). The realist synthesis type of systematic review (as developed by Pawson, 2002) extends the thinking behind policy theory-based approaches to synthesise high-quality evaluations that focus primarily on the mechanisms behind how and why interventions work (or fail) (Warren, 2014b). However, systematic reviews are resource-intensive and usually beyond the scope of policy evaluations sought by governments. Some governments, such as the UK, are increasingly using rapid evidence assessments (REA) to collect data on the impacts of policies (see UK Civil Service, 2014). Unlike systematic reviews, which often require a team of researchers working over a minimum of one year, a single evaluator can undertake a REA in less than six months. REA is a narrower and condensed version of a systematic review. Review methods are discussed in chapter three.

Evaluations of DSM policy are typically either ex-ante or ex-post though some governments, such as the USA, favour a combination of both evaluation types. Ex-ante appraisals look at the expected effects of a policy (for example, ‘deemed’ energy savings and engineering estimates), whereas ex-post evaluations look at the empirical results of a policy (for example, measuring the actual impacts from monitoring studies) (Fischer, 1995). Ex-ante appraisals can only estimate in advance the potential effectiveness of a policy but are much less resource-intensive to carry out (Warren, under review). In contrast, ex-post evaluations require measurements and monitoring of policy impacts to be conducted, which increases the reliability of the estimates of the impacts but are more resource-intensive to carry out (Warren, under review). Combined approaches are useful where ex-post evaluations feed into improving the accuracy of modelling tools in ex-ante appraisals (Mundaca and Neij, 2010). However, as Stern and Vantzis (2014) show, evaluators in North America tend to undertake ex-post evaluations, whereas in Europe, evaluators more commonly conduct ex-ante appraisals. Stern and Vantzis (2014) argue that Europe can learn from the North American experiences in conducting ex-post evaluations to improve the reliability of policy evaluations. In addition to the type of evaluation, who evaluates the policy is similarly important. Ex-post evaluations undertaken by governments can be politically damaging if they show a policy to have failed and it is possible that biases could exist in such evaluations. As a result, there is arguably a role for independent third parties to undertake evaluations of DSM policies using a combination of ex-ante and ex-post methods.

In calculating energy savings in all types of DSM policy evaluation, it is crucial to consider the factors summarised in figure 12 (the ‘perfect’ evaluation), which can be broken down into: calculating the energy consumption baseline, making gross savings adjustments, the attribution of energy savings, and monitoring and verification. The figure is adapted from the systematic review sample of high-quality DSM policy evaluations, which is discussed in chapter three. The energy consumption baseline refers to how the consumption that would have occurred without the programme (the baseline) was calculated. Making gross savings adjustments looks at how adjustments to energy savings were made, such as the lifetime of the measures installed or the degree of rebound effects, where financial savings from reduced energy consumption can be used to increase energy use overall, either directly (through the same activities) or indirectly (through other energy-consuming activities) (Sorrell, 2007a; Sorrell *et al.*, 2009). The attribution of energy savings refers to, for example, whether or not the programmes resulted in additional savings to what would have occurred had they not been implemented (additionality), or the degree of free ridership, where consumers would have still engaged with DSM in the absence of the programmes. Monitoring and verification looks at the degree of monitoring to assess the actual impacts of the programmes and to ensure that energy savings are not double counted.

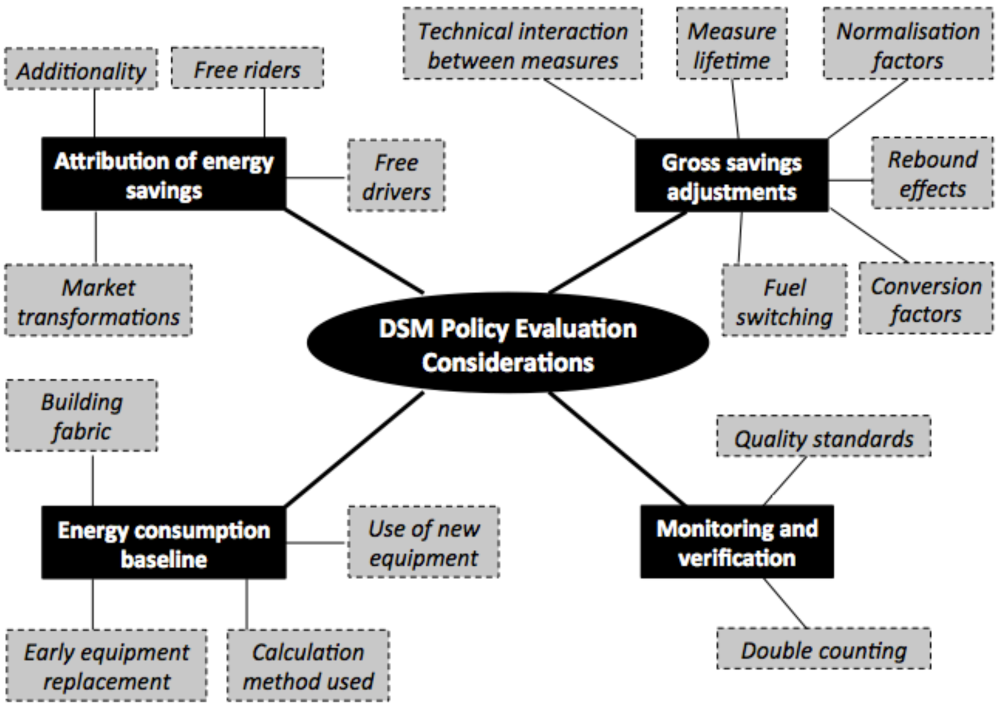


Figure 12: important considerations in DSM policy evaluation

2.4.2 Practice

In practice, DSM policies tend to be evaluated ex-ante with some small-scale ex-post monitoring to ensure that reality is still reflected. For example, the UK has used ex-ante approaches to evaluate its utility obligations since the first obligation began in 1994 (the *Energy Efficiency Standards of Performance* (EESoP 1)), which were based on the lifetime savings estimates from various installed measures. However, as Staniaszek and Lees (2012) argue, the accuracy of the evaluations has improved as energy savings were adjusted and validated over years of monitoring. The *Carbon Emissions Reduction Target* (CERT), which ran from 2008-2012, used a more combined approach with modelled data and data submitted by electricity suppliers (samples of which were independently verified and audited), alongside a programme of physical monitoring undertaken by the UK Energy Saving Trust (Staniaszek and Lees, 2012). Similar methods of evaluation are used in other European countries, such as Denmark and France, where standard values are often used and specific calculations are made where activities do not fit into particular standard categories.

Figure 12 conveys the theoretically ‘perfect’ policy evaluation, which in practice is rarely conducted due to resource constraints, such as funding, time, labour, expertise, practical difficulties and ethical considerations. Thus, evaluators make trade-offs between what is accounted for and what is not. Despite this, there is a clear pattern as to what evaluators focus on, as shown in the systematic review sample of high-quality DSM policy evaluations (discussed in chapter three). For attribution of energy savings, additionality is frequently discussed, followed by free riders. For gross savings adjustments, measure lifetimes are usually included but rebound effects are rarely calculated. However, rebound effects are difficult to determine and there are still large uncertainties over their true extent (see Sorrell, 2007a and Sorrell *et al*., 2009 for comprehensive reviews). The energy consumption baseline is arguably the most important aspect of DSM policy evaluation (Vreuls, 2014). It is a challenge to calculate due to the uncertainties in determining how consumers would have behaved without the DSM programme. However, extrapolating from past demand patterns is usually used to do this.

2.4.3 Evidence Base

The quick scoping review and the systematic review (outlined in chapter three) highlight that the evidence base for high-quality DSM policy evaluation is reasonably limited, with the majority of evaluations having been conducted by industry, rather than by academia or by governments. Chapter three discusses evidence quality and chapter four analyses the current evidence base for DSM policy evaluation. As highlighted in chapter one and further discussed in chapter three, there has been a wealth of literature on DSM since the energy crises of the 1970s. However, a large proportion of the work has focused on utility programmes (independent of government) in the USA, which were introduced primarily for portfolio and load management purposes. Outside of the USA, DSM research has examined technological (or behavioural) trials or modelled the future potential of DSM in the energy system. Thus, despite much research on DSM, there is limited analysis from academia and governments on the policy side of DSM. This forms part of the justification for the research and this thesis aimed to capture and analyse a large proportion of the high-quality evidence base that exists for DSM policy evaluation.

2.5 Summary

A review of the definitions of demand-side management (DSM) since the energy crises of the 1970s, when DSM was first implemented in national policy, reveals a shift from traditional discussions that focussed primarily on utility load management activities to broader definitions that encompass government policy for the full range of demand-side activities. Thus, DSM in the 21st century refers to technologies, actions and programmes on the demand-side of energy meters that seek to manage or decrease energy consumption, in order to reduce total energy system expenditures or contribute to the achievement of policy objectives such as emissions reduction or balancing supply and demand.

Some of the primary drivers for demand-side activities outside of government regulation are for utilities to defer investment in generation infrastructure and to encourage more active market participation from consumers in order to contribute to balancing supply and demand at least-cost. However, the business model of traditional utilities is based on increasing profits through selling more energy, and thus there is a disincentive for utilities to encourage consumers to use less energy. Furthermore, research from the behavioural economics and social psychology literature has shown that consumers exhibit ‘wider rationalities’ where they are influenced by social norms, habits, peers, competitors, family and neighbours, which have led to consumers’ limited engagement with DSM historically.

Business and policy models to overcome the disincentive for utilities to engage with DSM have been implemented at a state-level in half of the states in the USA. Alternative utility business models (UBM), such as decoupling and lost revenue adjustment mechanisms (LRAM), appear to have worked well in markets with public utilities that are economically regulated by public utilities commissions. There is potential for such models to be transferred to other countries with similar energy market structures. However, UBM would be challenging to implement in fully liberalised markets, such as those in the European Union (EU), due to the lack of required economic regulation – instead, prices are determined based on competition.

DSM policy can be categorised into policies that are market-based, regulatory, fiscal, information-based or voluntary. UBM is a set of policies that come under the fiscal category. Evaluations of DSM policy tend to be ex-ante appraisals (undertaken before a policy has been implemented), ex-post evaluations (undertaken after a policy has been implemented) or a combination of both types. Rigorous policy evaluation should include a critical appraisal of the full policy process from policy proposal through design and implementation to the post-policy evaluation stage. However, in practice this is rarely done, as chapter four discusses. Important considerations in the evaluation of DSM policy are the attribution of energy savings (such as additionality), making gross savings adjustments (e.g. accounting for rebound effects), calculating the energy consumption baseline, and undertaking monitoring and verification. DSM policy evaluation is the focus of this research and is discussed in more depth in the next chapter.

3 Chapter 3: Research Design

3.1 Research Focus

This chapter discusses and justifies the methodological approach employed in the research. Section 3.1 outlines the research focus in terms of research aims and questions; section 3.2 discusses the overarching philosophy and methodology of the thesis; section 3.3 details the methods used for data collection; section 3.4 discusses and justifies the primary method of the research (systematic reviews); and section 3.5 details the techniques used for data analysis and synthesis.

3.1.1 Research Aims

Chapter one gave an overview of the main aim of the research: to determine the mechanisms behind DSM policy success and failure.

The research aims to explore the key factors for success and failure for different types of DSM policy, such as those shown in figure 9 in chapter two. The literature review (which covered the 389 academic, industrial and policy documents included in the bibliography) highlighted that the evidence base for DSM policy evaluation is reasonably limited, with the focus of previous DSM research primarily focussing on trials and pilots of DSM technologies (such as smart meters and time-of-use tariffs), modelling studies of the future potential of DSM in the energy system (such as the future contribution of heat pumps and demand response) and programmes undertaken by utilities independent of government regulations (primarily for load management purposes). Research that has concentrated on the policy side of DSM has been limited to impact assessments conducted before the policy has been implemented (as is common in European evaluations) or after the policy period has ended (as is common in North American evaluations). However, the main research gap is the lack of attention given to the mechanisms behind how and why a policy succeeded or failed, as opposed to simply identifying its impacts.

3.1.2 Research Questions

In order to meet the research aim, the following research questions were developed:

1) What DSM policies have been implemented around the world with high quality documented evaluations?
2) How and why do DSM policies succeed or fail, and what policies have been successful?

To answer research question one, the thesis examines the evidence base for DSM policy evaluation and assesses the quality of the evidence in order to map out the global distribution of DSM policy implementation. The focus on high quality evaluations not only narrows the scope of the research to a project feasible within the timescales of a PhD but also overcomes issues of identifying countries/states that have implemented DSM policies but did not evaluate them, or evaluated the policies but did not translate them into English. This is discussed further in section 3.4. Answering the research question also allows the identification of the types of DSM policy that are more commonly implemented and in which countries/states. Furthermore, the global spatial distribution of the high quality evidence base can be determined, as well as temporal patterns showing how the evidence base has changed over time. However, temporal analysis is not an essential part of answering research question one and it is discussed as an area for further research in chapter six. Chapter four is the results and discussion chapter for research question one.

To answer research question two, the thesis determines the mechanisms behind DSM policy success and failure. The analysis identifies the key factors for success and failure for different types of DSM policy and for different countries/states, and identifies which DSM policies have been successful and unsuccessful in particular countries/states. Policy success is defined and justified in sub-section 3.5.2 and in chapter five. Chapter five is the results and discussion chapter for research question two.

3.2 Methodology

3.2.1 Research Philosophy and Approach

The research questions are more focussed on exploration and theory building rather than theory testing by accepting or rejecting hypotheses. The thesis employs the *research process onion* concept for research design, which was developed by Saunders *et al*. (2007), as it provides an excellent way to visualise and categorise the scientific path from research philosophy to specific methods. Saunders *et al*. (2007) split the *research process onion* into six stages: ‘philosophy’, ‘approach’, ‘strategy’, ‘choice’, ‘time horizon’ and ‘techniques’. The concept is visualised in figure 13 and is labelled to show the research design of the thesis.

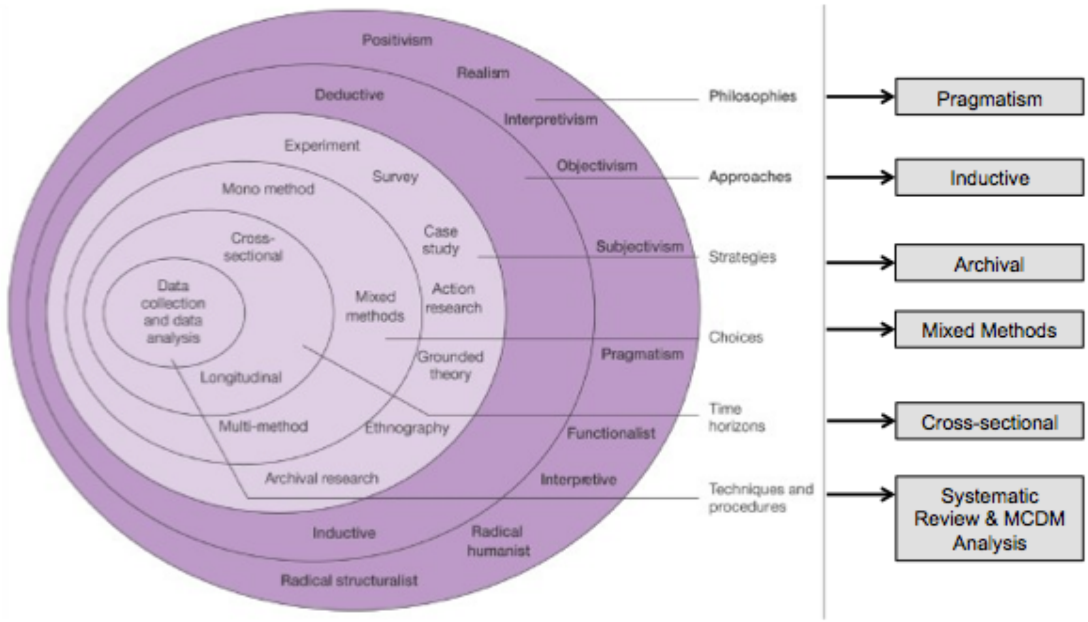


Figure 13: the research design of the thesis

In the outermost sixth layer, the thesis fits into the pragmatism research philosophy, which focuses on the method and using methods that are well suited to meeting the research aims (Morgan, 2007). This is in contrast to philosophical paradigms that give less freedom in the methods and tools that could be employed and instead focus on the approaches and strategies that should be employed, as can be the case with positivism and interpretivism (Saunders *et al*., 2007). In the fifth layer, the 'approach' employed in the thesis is inductive as data is synthesised and theories are built from the data rather than tested. The alternative research 'approach' is deductive, which involves the development of theory from the literature, which is then tested by accepting or rejecting hypotheses (Kellstedt and Whitten, 2013). However, as previously argued, the inductive approach is more suitable for research that is exploratory in nature.

3.2.2 Research Strategy, Choice and Time Horizon

The fourth layer of the *research process onion* identifies the research 'strategies' that are employed. Saunders *et al*. (2007) identify seven main 'strategies': experiments, surveys, case studies, action research, grounded theory, ethnography and archival research. The thesis uses archival research (in its broadest sense) as it primarily focuses on high quality documented evaluations of DSM policy. In the third layer, Saunders *et al*. (2007) state three research 'choices': mono-method, multi-methods and mixed methods. Mono-method refers to a single method that is employed, which can be either quantitative or qualitative. Multi-methods are an extension of this where two or more methods are used that are either quantitative or qualitative (but not both). In contrast, mixed methods refer to the use of two or more methods that are quantitative and qualitative. Saunders *et al*. (2009, 5th edition) break down the mixed methods category further into 'mixed-method' and 'mixed-model'. The former includes quantitative and qualitative methods but they are not combined in analysis procedures, whereas the latter includes quantitative and qualitative methods that are combined in analysis procedures. Although Saunders *et al*. (2009) provide useful methodological categories, the mixed methods category fails to categorise those methods that have both quantitative and qualitative aspects to them (such as certain types of interviews (discussed in sub-section 3.4.8) and surveys). It is likely that these would be included within the 'mixed-model' category due to the lack of separation of the quantitative and qualitative aspects of the method(s). If two or more of such methods were used, it is possible that a new category is needed, although an extension of the 'mixed-model' category is arguably more appropriate.

In the second layer, the research ‘time horizon’, Saunders *et al*. (2007)’s two options, cross-sectional and longitudinal, are arguably limited. Cross-sectional studies undertake spatial analyses, whereas longitudinal studies undertake temporal analyses. However, there are a number of studies that undertake panel data analyses, which are multi-dimensional and cover both spatial and temporal aspects (Hsiao, 2003). This is missing from the *research process onion* concept. Despite this limitation, the concept is still a useful visualisation of the different approaches to science. In the innermost first layer, the research ‘techniques and procedures’ refer to the specific methods for data collection and data analysis, such as document analysis, interviews, focus groups, modelling (such as energy system modelling and economic modelling), controlled experiments, physical monitoring, etc. Section 3.3 discusses the methods used in this research.

3.3 Methods

3.3.1 Methods for Policy Analysis

Section 3.2 established the research methodology in order to provide the most appropriate research design for answering the two research questions. However, once the five outer layers of the *research process onion* concept have been established, the most appropriate methods for answering the research questions need to be determined. The focus of the thesis on policy analysis and the broad and global nature of the research questions requires a method that can comprehensively evaluate the evidence base at the level of national or sub-national state policy.

Yang (2007) states that “policy analysis involves using quantitative and/or qualitative techniques to define a policy problem, demonstrate its impacts and present potential solutions” (p. 349). Although this is a common definition of policy analysis, it is narrow in its approach, as it focuses only on policy impacts rather than policy mechanisms, which are crucial in the identification of improvements for future policies. However, the definition does highlight that there are numerous methods for policy analysis, which can be quantitative, qualitative or mixed methods. Quantitative techniques include: modelling, quantification of inputs and outputs, descriptive statistics, statistical inference, operations research, cost-benefit analysis, risk-benefit analysis, Q methodology, multi-criteria decision-making analysis, cost-effectiveness analysis, survey research and impact assessments (Yang, 2007, p. 349; Nagel, 2002, p. 139). Qualitative techniques include: focus groups, interviews, participant observation, ethnographic research, action research, content analysis, document analysis, case study analysis and grounded research (Yanow, 2007, p. 405; Sadovnik, 2007, p. 421). Mixed methods include a combination of quantitative and qualitative methods, as discussed in sub-section 3.2.2.

Bardach (2005) argues that the following five aspects are crucial for undertaking ‘best practice’ in policy analysis:

- Developing realistic expectations
- Analysing ‘smart’ practices
- Observing the practice (functions and features)
- Describing generic vulnerabilities
- Assessing transferability

The framework is useful and appropriate for analysing the research questions, as the research boundaries (discussed in section 3.4) allow the development of realistic expectations of what can be covered within the timescales of a PhD; ‘smart’ or successful DSM policies are identified; the mechanisms (functions and features) of DSM policies are observed; failure factors for DSM policies are identified and described (vulnerabilities); and the transferability of successful DSM policies is highlighted as an area for further research. However, the challenge is the identification of one or more methods that can cover all five aspects of Bardach (2005)’s ‘best practices’ framework. This is discussed in the following sub-sections. However, in summary, review methods are particularly effective at capturing large amounts of data to answer research questions that look at national and international scales. They can be qualitative, quantitative or mixed methods, and in contrast to literature reviews, they aim to collect data for analysis (Warren, 2014b).

3.3.2 Research Methods

The research utilises review methods in order to answer the two research questions. There are five main types of review method: quick scoping reviews, rapid evidence assessments (REA), systematic reviews, multi-arm systematic reviews and review of reviews. In practice, the five review methods can be combined into two main types: systematic reviews (covering multi-arm systematic reviews as well) and REAs (covering review of reviews and quick scoping reviews as well), which differ in quality, detail and the time and resources needed to undertake them. Figure 14 summarises the key differences between them. It is adapted from the UK Civil Service (2014), which provides a good overview of the main types of review methods. The most appropriate review method for comprehensively answering the research questions within the resources available is a systematic review, which is discussed and justified in section 3.4.

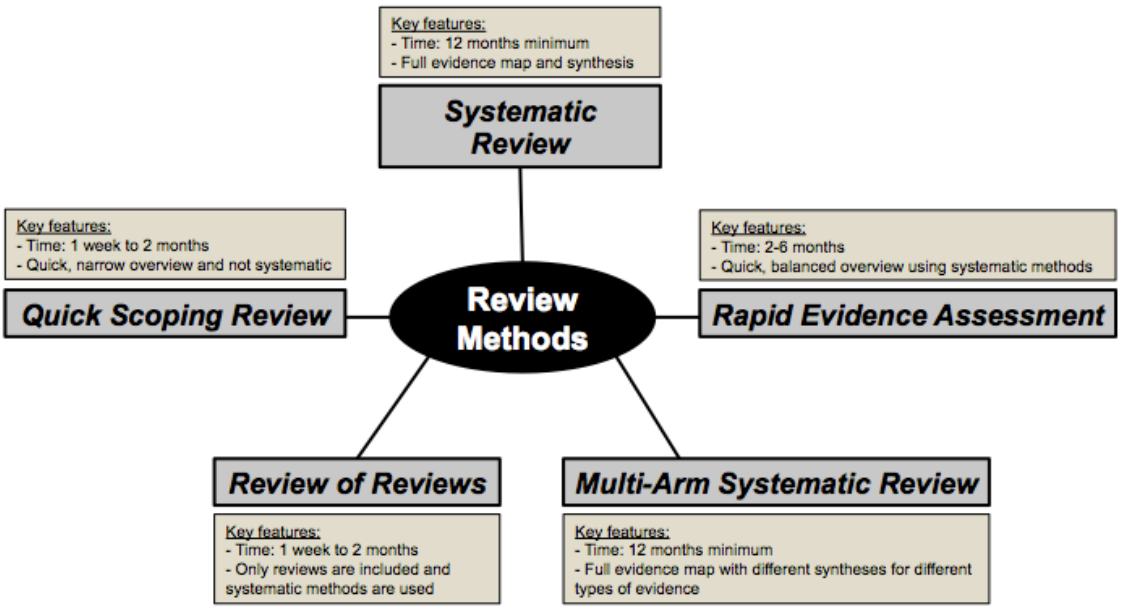


Figure 14: a comparison of the main types of review methods

Quick scoping reviews do not use systematic techniques (such as detailed search strategies, inclusion criteria and data quality assessments – discussed in section 3.4), and thus have a constrained focus and should be used to obtain a general overview of studies conducted on a specific and narrow topic (UK Civil

Service, 2014). Although they are the lowest quality of review method, their strength lies in the limited resources required to undertake them, such as time (usually one week to two months), funding, sources (and access to them) and labour (UK Civil Service, 2014). Quick scoping reviews are also essential parts of conducting systematic reviews and REAs, as they must be undertaken first in order to develop the review protocol for systematic reviews and REAs, as discussed in section 3.4.

Unlike quick scoping reviews, REAs comprehensively cover a specific topic by using systematic techniques. However, the focus of REAs must be narrow in order to undertake them in less than half the time it takes to conduct a systematic review (2-6 months) (Ganann *et al*., 2010). REAs can be viewed as shorter versions of systematic reviews as they are conducted in the same way, but require fewer resources, such as time, funding, sources and labour. Thus, they are particularly useful for policy makers that require synthesis of knowledge in less than six months (Ganann *et al*., 2010).

Systematic reviews aim to provide a full evidence map and synthesis of existing research (UK Civil Service, 2014). It is thus difficult to conduct a systematic review in less than one year full time (Warren, 2014b). As such, they are resource-intensive and require time for obtaining the relevant training in order to learn how to conduct them (Gough *et al*., 2012). However, where resources allow, they are often considered to form the top of the *Hierarchy of Evidence*, particularly in the medical sciences, as discussed in section 3.4

Multi-arm systematic reviews follow the same principles as systematic reviews except that different types of systematic review (discussed in section 3.4) are conducted for different types of evidence (UK Civil Service, 2014). For example, meta-analyses are conducted where quantitative data from monitoring and trials are available, which can then be combined with syntheses of data from qualitative studies, such as interviews or focus groups. Such reviews are the most resource-intensive to carry out and usually require a team of researchers working for a minimum of one year (UK Civil Service, 2014).

Review of reviews use systematic techniques and normally take one week to two months to carry out. They require fewer resources than systematic reviews and REAs, such as time, funding, sources and labour, but require more resources than quick scoping reviews. A further strength of the method is that it focuses only on reviews of the topic in question (UK Civil Service, 2014), which helps to narrow the scope for the inclusion criteria and saves time by not needing to collate primary studies. However, this is also the main weakness of review of reviews, as each review varies in its quality and scope and would need to be assessed for its use of systematic methods. Smith *et al*. (2011) argue that the quality of the reviewed evidence base in each review needs to be described and summarised, and the strength of each review’s conclusions should be discussed.

The five main review methods are classified as methods, as they aim to collect data and evidence for analysis. In contrast, literature reviews aim to collate relevant studies on a particular topic and appraise them in order to draw conclusions. The process includes the identification of the key arguments and research gaps without using systematic techniques. Thus, the purpose of literature reviews is not to collect data for analysis and hence they are not classified as a review method.

3.4 Systematic Review

3.4.1 Improving Evidence Quality

Systematic reviews involve collating and synthesising all of the work that has been done on a particular intervention, trial or programme to better understand what works and what does not (Petticrew and Roberts, 2006). A number of disciplines, such as the medical sciences, use the *Hierarchy of Evidence* to determine the quality of research studies. As shown in figure 15, developed and described by the University of Illinois, editorials and expert opinion form the base of the pyramid, with case series and case reports above this, case-control studies above this, cohort studies in the middle, Randomised Control Trials (RCTs) second from top, and systematic reviews at the top of the pyramid.

Figure 15: the *Hierarchy of Evidence*

Expert opinion refers to the judgements of defined experts in the field, as collected in interviews or focus groups. Case reports are a type of anecdotal evidence and tend to be less scientifically rigorous than study types above it due to associated bias, lack of random sampling and the absence of control groups (Daramola and Rhee, 2011). However, some authors in the medical sciences have argued that they still have an important role to play in research, as they have a high sensitivity for detecting novelty and providing new ideas (Vandenbroucke, 2001). Case-control studies are a type of observational and retrospective study in which different groups of subjects are observed to infer causal attributes rather than placed in controlled trials (University of Illinois, 2013). Cohort studies are a type of longitudinal study that analyse risk factors in cohorts of people with the same characteristics, such as age, and observes them to see if a particular condition being researched develops (University of Illinois, 2013). RCTs involve randomly allocating subjects to receive (treatment group) or not receive (control group) an intervention and keeping all other conditions the same for all participants (Akobeng, 2005).

The *Hierarchy of Evidence* is relevant in the medical sciences, particularly for the testing of new treatments, such as new drugs. However, the hierarchy is less relevant for research problems where study types towards the top of the pyramid, such as RCTs, cannot be conducted for practical, economic, or ethical reasons, for example, nationally testing the influence of school teaching quality on entrance levels to university. In an RCT, all pupils in a country would need to be randomly and blindly assigned to under-achieving schools and high-achieving schools regardless of background and locality, and subjected to exactly the same external conditions other than teaching quality (Warren, 2014b). For a host of practical, economic and ethical reasons this could not be conducted. In the energy policy field, RCTs would similarly be less appropriate for the analysis of, for example, national policies to reduce energy consumption, as the two primary strategies to do this would be to compare different countries (or states) implementing the same policy (e.g. subsidies) or to implement different policies (e.g. subsidies and performance standards) within the same country (or sub-national state) under exactly the same conditions. In the case of the former, policy types, designs, implementation processes, evaluation processes and culture may greatly differ between countries (or states), thus making it difficult to account for all confounding variables (Kellstedt and Whitten, 2013). In the case of the latter, control groups cannot be established for obvious practical reasons. However, RCTs in the energy field have been limited to date and there is a clear need to undertake them on more practical scales, such as small-scale trials. There is some evidence to suggest that this is beginning to change and recent examples in the UK include an RCT of residential demand response (time-of-use tariffs) (Nicolson, 2014) and the DECC-John Lewis energy labelling trial (inclusion of long-term running costs on labels). Nevertheless, at the national policy scale, the challenges of conducting randomised and double-blinded RCTs are apparent.

The University of Illinois's *Hierarchy of Evidence* is characteristic of many hierarchies in primarily quantitative disciplines and is useful where research questions can more effectively be answered by quantitative analysis. However, there are a number of research problems where this would be less appropriate and would be better solved through qualitative studies, using methods such as interviews, focus groups and participant observation. Some researchers, such as Daly *et al*. (2007), who also work within the medical sciences, have proposed an alternative hierarchy, which is designed specifically for assessing qualitative research. The hierarchical pyramid has generalisable studies at the top as 'Level I', conceptual studies underneath this as 'Level II', descriptive studies underneath this as 'Level III' and single case studies forming the base as 'Level IV'. Unlike the University of Illinois's hierarchy, Daly *et al*. (2007)'s hierarchy is not structured by specific method but by feature, which is arguably more appropriate for qualitative research.

Although systematic reviews are well established in the medical sciences, particularly through the Cochrane Collaboration (providing a database of >5,000 systematic reviews), it has had limited attention in other disciplines. However, the Campbell Collaboration, established in 2000, is beginning to apply the method to other policy areas, such as education, crime and justice, and social welfare. Nevertheless, it has had limited attention in the energy policy field and there have been calls for this to be undertaken in order to improve the quality of the evidence-base for informing energy policy (Sorrell, 2007b; Warren, 2014b). The Collaboration for Environmental Evidence (CEE) is a collaboration that may go someway to addressing this issue. CEE gathers systematic reviews and evidence maps in the environmental field, though at the time of writing only two have been conducted in the energy field and stored in the CEE database (Papathanasopoulou *et al*., forthcoming (impacts of energy systems on marine ecosystems), and Watson *et al*., 2012 (barriers to energy services in the poorest countries)).

Over the course of the PhD, the application of systematic reviews has begun to develop in other parts of the energy field (independent of CEE), such as bioenergy research (e.g. Muench and Guenther, 2013; Gurwick *et al*., 2013; Rehfuess *et al*., 2014) and climate change research (e.g. Porter *et al*., 2014). In 2014, a PhD thesis was published that conducted a meta-evaluation of energy efficiency evaluations, though this did not focus on policy and instead examined the practice of evaluation rather than the use of systematic reviews (Brown, 2014).

Systematic reviews are an effective method for collating the results of all of the previous studies that have been conducted on a given programme or intervention. Unlike traditional literature reviews, which simply summarise what is known on a topic without explaining the criteria used to identify, include and methodological appraise relevant studies, systematic reviews employ systematic and replicable techniques to improve the quality of the evidence base. As Petticrew and Roberts (2006) argue, traditional reviews often have poor specification of the review topic, use evidence selectively and opportunistically, give limited attention to methodological quality and lack transparency. Sorrell (2007b) highlights that systematic reviews aim to overcome these issues, though he is sceptical about its potential to have a comparable impact in energy policy as it has done in healthcare. Sorrell (2007b) argues that systematic reviews still tend to answer micro questions regarding technical efficiency rather than macro policy questions. This thesis aims to contribute to filling this methodological research gap.

Documents that may be included in a systematic review are published and unpublished material, academic and 'grey' literature (such as policy documents, industrial publications and consultancy reports), and peer reviewed and non-peer reviewed documents. Furthermore, these documents may be available online or obtained through hand searching, and may or may not be available in English. However, as explained in the following sub-sections (3.4.3 to 3.4.6), all relevant documents are subjected to specific inclusion criteria and data quality assessments before the final sample is established.

3.4.2 Systematic Reviews: Background and Types

There are a number of different types of systematic review each with varying methods of conduction, which can be classified under three main review types: interpretive (e.g. qualitative synthesis), integrative (e.g. meta-analysis) and mixed methods (e.g. realist synthesis). Table 4 summarises the different types of systematic review.

Category	Systematic Review Type
Interpretive	Narrative Summaries
	Thematic Summaries
	Grounded Theory
	Meta-Ethnography
Integrative	Content Analysis
	Case Survey
	Qualitative Comparative Analysis Method
	Meta-Analysis
Mixed Methods	Meta-Study
	Realist Synthesis
	Miles and Huberman's Cross-Case Techniques
	Framework Synthesis
	Thematic Synthesis

Table 4: the different types of systematic review

Interpretive systematic reviews are primarily qualitative, though they can have some quantitative aspects. Examples include: narrative summaries, thematic summaries, grounded theory and meta-ethnography (Dixon-Woods *et al*., 2005; Snilstveit *et al*., 2012). In contrast, integrative systematic reviews are primarily quantitative, though they can have some qualitative aspects. Examples include: content analysis, case survey, qualitative comparative analysis method and (Bayesian) meta-analysis (Dixon-Woods *et al*., 2005; Snilstveit *et al*., 2012). Mixed methods systematic reviews aim to combine the benefits of both qualitative and quantitative approaches in order to reduce the drawbacks of individual approaches. Examples include: meta-study, realist synthesis, Miles and Huberman's cross-case techniques, framework synthesis and thematic synthesis (Dixon-Woods *et al*., 2005; Snilstveit *et al*., 2012). Dixon-Woods *et al*. (2005) and Snilstveit *et al*. (2012) provide excellent summaries of the methodological details, strengths and weaknesses of the different individual types of systematic review and thus are not repeated here. However, a comparison of the three broad categories of systematic review is provided.

The interpretive approach synthesises the concepts identified in the primary studies into a higher-order theoretical structure (Dixon-Woods *et al*., 2005), whereas the integrative approach is concerned with amalgamating data for statistical analysis (Dixon-Woods *et al*., 2005; Harden and Thomas, 2005). In the context of this research, the former approach is particularly useful for analysing contextual factors within which a policy is implemented. However, it is limited with respect to analysing policy impacts and the development of hierarchies of effectiveness (Snilstveit *et al*., 2012). The integrative approach overcomes this limitation and can analyse policy impacts and develop hierarchies of success. It is one of the most widely used forms of systematic review and is usually the approach adopted in the medical sciences. Nevertheless, Pawson (2002a) argues that meta-analyses reduce each individual programme to a single measure of effectiveness, which is put into an aggregate measure for that sub-class of programmes. This results in important explanatory content being removed (Pawson, 2002a). The strength of the interpretive approach is the weakness of the integrative approach, as the latter does not analyse the contextual factors that may impact on the transferability of programmes or policies between places.

Both broad categories of systematic review (interpretive and integrative) generally perform poorly in determining the mechanisms behind how and why particular programmes or interventions work. As Pawson (2006) notes:

"…in order to identify causal connections, we need to understand outcome patterns rather than seek outcome regularities,,,We rely on mechanisms to tell us why interconnections occur…The mechanism explains what it is about the system that makes things happen." (pp. 22-23)

Pawson (2002b) proposes a mixed methods approach, realist synthesis, as a solution for combining the benefits of different approaches while reducing their drawbacks. He describes realist synthesis as theory-driven and focused on the underlying programme theory and mechanisms driving an intervention. The method is particularly effective at understanding how and why a particular programme (or policy) works by looking at its underlying mechanisms, thus extending the boundaries of what qualitative syntheses can do to better assess transferability between places, and moving the attention away from looking solely at specific outcomes rather than context, as is the focus of meta-analyses.

Despite these benefits, Dixon-Woods *et al*. (2005) highlight that the realist synthesis approach tends to treat all forms of evidence as equally authoritative and there is a lack of explicit methodological guidance on how to conduct such analyses in practice. Nevertheless, realist synthesis allows a comprehensive analysis of policy impacts where policies were designed, implemented and evaluated differently by different groups in different places, by looking at what works and why it works. As such, realist synthesis is the most appropriate type of systematic review in order to answer the three research questions. In the following sub-sections (3.4.3 to 3.4.6), the thesis shows how the two main criticisms of realist synthesis (the homogenous treatment of evidence and a lack of practical guidelines for conducting it) can be overcome. Thus, the research extends the work of Pawson (2002b), as well as applying realist synthesis to the energy policy field (Pawson, 2002b discusses the method in relation to social care).

Systematic reviews are a useful method for collecting all of the relevant and current research undertaken on a given topic. Although there is no limit on the temporal extent of documents included in the review, the most important aspect is to collect the most robust evidence on the topic. However, this leads to the main limitations of systematic reviews. They are time-consuming and resource-intensive, usually carried out by a team of researchers over a number of years and often involve a learning curve to develop the required review skills, especially if the method is being used in a field where it has had limited application. Although the available resources determine the breadth and scope of a review, in general, the societal funding of reviews is minimal compared with primary research (Gough *et al*., 2012, p. 11). Furthermore, as Gough *et al*. (2012) argue, in determining what documents are included in the review, researchers working within different research paradigms (and with potentially different perspectives) may vary in what they deem to be a good quality study (p. 13). Thus, the inclusion criteria and study quality assessment need to be clearly defined, justified, and replicable.

Undertaking a pilot study can test how robust the systematic review protocol is. This involves undertaking a much smaller version of the systematic review, using one or two search terms in a small number of specified databases, where two or more researchers carry out the pilot study following exactly the same methodological protocol. The process for collecting and synthesising the evidence would be compared between researchers and any differences flagged up in order to tighten up the protocol until it is more replicable. Sub-section 3.4.7 outlines the pilot studies that were undertaken in the strengthening of the systematic review protocol developed for this research. The systematic review protocol is detailed and justified in the following sub-sections (3.4.3 to 3.4.7).

3.4.3 Systematic Review Protocol: Stages 1-2

There are eight main stages in conducting systematic reviews, as shown below (adapted from Harden and Thomas, 2005):

1. Review questions and boundaries
2. Selection of systematic review type
3. Comprehensive search
4. Inclusion criteria
5. Quality assessment
6. Data/information extraction
7. Synthesis of findings
8. Dissemination of findings

The subsequent sub-sections follow the above, adapted version of Harden and Thomas (2005)'s eight-stage framework to detail the systematic review protocol employed.

For stage one of the review protocol, the research aims to answer the following research questions, as detailed earlier in the chapter:

1. What DSM policies have been implemented around the world with high quality documented evaluations?
2. How and why do DSM policies succeed or fail, and what policies have been successful?

For stage two of the review protocol, the research employs the realist synthesis type of systematic review, as the method's primary strength is in understanding how and why particular programmes (or policies) work by examining the underlying mechanisms that determine whether they succeed or fail (Pawson, 2002a). The three main aspects of the analysis are to assess the patterns of implementation (qualitatively and quantitatively), extract the policy impacts (quantitatively) and determine the policy mechanisms (qualitatively) in order to answer the research questions.

3.4.4 Systematic Review Protocol: Stages 3-4

For stage three of the review protocol, the comprehensive search involves the use of pre-determined databases in order to locate relevant DSM policy evaluation documents. A quick scoping review was conducted to identify the key databases and institutions that publish documents on DSM. Quick scoping reviews do not use systematic techniques and the broad phrase, 'demand-side management', was inputted into the three largest academic databases (Web of Science, Scopus and Google Scholar) in order to identify the academic journals that are more prominent in publishing DSM research. It is important to note that at this stage the search did not aim to focus on DSM policy, but DSM in general, to ensure that potentially key databases were not excluded before the systematic review search was conducted. From the quick scoping review, the journals listed below feature prominently in publishing DSM research:

- Energy Efficiency
- The Electricity Journal
- Energy
- Energy Policy
- Energy Economics
- Energy and Buildings
- Resource and Energy Economics (REE)
- The Energy Journal
- Electric Power Systems Research (EPSR)

The primary reason for using specific journal databases within the three larger databases that cover them was to ensure that a more comprehensive search could then be conducted. The quick scoping review identified papers within the journals that did not appear in the Web of Science/Scopus/Google Scholar searches. This led to the decision to increase the breadth of the databases that would be included, which was confirmed as an appropriate strategy by a panel of five academics and by a systematic review expert.

For non-academic literature (‘grey’ literature), few specific databases exist. From the MCDM expert interviews (discussed in sub-section 3.4.8), some participants gave referrals on databases and websites that could be used:

- Open Grey (a database of European grey literature, formerly called SIGLE – ‘System for Information on Grey Literature’)
- National Grid (using a search within the main website)
- US Federal Energy Regulatory Commission (FERC) (using a search within the main website)
- US Department of Energy (DoE)’s Energy Citations Database (since the systematic review was conducted, the original website has now changed to the more comprehensive: http://www.osti.gov/scitech/)
- UK National Audit Office (NAO) (using a search within the main website)
- UK Public Accounts Committee (PAC) (using a search within the main website – however, since the systematic review was conducted, the original website has now changed and this can no longer be done)

Google searches were also conducted to find other grey literature, such as consultancy reports, institutional research reports, government impact assessments and policy evaluations, PhD theses and conference proceedings. Although 14.8 million hits were returned from using the search term, ‘demand-side management’, in Google, the first ten pages were deemed comprehensive enough for the identification of the key institutions, government departments, consultancies, industry groups and non-governmental organisations that publish documents on DSM. It is important to reiterate that the purpose of conducting a quick scoping review prior to the systematic review was to identify the sources of documents rather than the documents themselves. From the Google search, the organisations, institutions and government departments listed below were identified:

- Institution databases:
    - American Council for an Energy-Efficient Economy (ACEEE)
    - European Council for an Energy-Efficient Economy (ECEEE)
    - International Energy Program Evaluation Conference (IEPEC)
    - International Energy Agency (IEA) DSM Programme
    - Institute of Electrical and Electronics Engineers (IEEE *Xplore* digital library)
- Institution websites:
    - International Partnership for Energy Efficiency Cooperation (IPEEC)
    - British Institute for Energy Economics (BIEE)
    - International Association for Energy Economics (IAEE)
- Industry group websites:
    - Association for the Conservation of Energy (ACE)
- Government and regulator websites:
    - US Department of Energy (DoE)
    - US Energy Information Administration (EIA)
    - UK Department of Energy and Climate Change (DECC)
    - UK Office of Gas and Electricity Markets (Ofgem)
    - Chinese National Development and Reform Commission (NRDC)
    - Australian Department of Industry (now called the Australian Department of Industry and Science)
    - Australian Energy Regulator (AER)
    - California Public Utilities Commission (CPUC)
    - European Commission Department of Energy

The quick scoping review thus identified the above 33 academic (9), industrial (11) and policy (13) databases and websites that have a track record in publishing documents on DSM. In addition to this, a database of DSM policy experts was integrated into the systematic review, which is discussed in sub-section 3.4.8. The remainder of the review protocol refers to ‘databases’, which includes both databases of documents and websites of relevant organisations.

Following the quick scoping review, the comprehensive search strategy was developed. Systematic reviews are usually undertaken by a team of researchers over a number of years, but due to the time and resource constraints of a PhD, the review had to be limited to one researcher in one year. As a result, the following review boundaries were established:

- One-year for data collection
- One researcher undertaking the data collection and analysis
- Peer review of the review protocol by a panel of academics
- Two pilot studies undertaken with two other researchers to check the robustness and replicability of the review
- The use of one search term: '"demand-side management" AND policies AND programmes"' in the pre-determined databases

The systematic review protocol was agreed by a panel of five academics (that are experts in DSM or policy analysis) and by one systematic review expert from the medical sciences (who has expertise in a range of different systematic review types, including those used in the social sciences).

As the systematic review could not be conducted in parallel by a second researcher, two pilots studies were undertaken with two other researchers to ensure that the review could be replicated and easily updated. This improved the robustness of the systematic review under the resource constraints of the PhD. The two other researchers received a concise version of the review protocol, which outlined the specific process for conducting the search, producing the final sample of documents, and extracting relevant data and information. The document is included in the Appendix (Appendix Figure 1) and the details of the pilot studies are discussed in sub-section 3.4.7. Following the pilot studies, the final review protocol was developed and subjected to further peer review through the submission and subsequent publication of a paper in the academic journal, *Energy Efficiency* (in Warren, 2014b).

Stage three of the review protocol involves the development of search terms in addition to the identification of databases. Under the resource and time constraints of the PhD, a trade-off between the breadth of databases covered and the number of search terms used in each database was made. As discussed above, just using the three large academic databases (and consequently increasing the number of search terms that could be used) has the potential to exclude papers that might not be retrieved unless an individual search of the specific journal databases is also conducted. Furthermore, Web of Science, Scopus and Google Scholar do not cover industrial and governmental literature. As the quick scoping review identified, the majority of DSM policy evaluations have been conducted in industry by institutions, non-governmental organisations and consultancies. Thus, there are knowledge gaps in the academic (and governmental) literature. In order to cover all 33 databases listed above within the one-year timeframe of the review, the number of search terms that could be used had to be reduced. The ten search terms given in table 5 were developed and tested in all 33 of the databases to evaluate the relevance of the documents retrieved and to refine them for improvement.

Search Terms (with and without Boolean logic)
"Demand-side management" AND programmes AND policies
"Demand-side management" AND programmes AND policy
"Demand-side management" AND program* AND polic*
("Demand-side management" AND programmes AND policies) AND (tariffs AND interruptible/curtailment OR "dynamic peak pricing" OR "critical peak pricing" OR "off-peak" OR "demand bidding" OR time-of-use OR real-time)
("Demand-side management" AND programmes AND policies) AND (labelling OR “performance standards" OR "utility obligations" OR "supplier obligations" OR "market transformation" OR "smart meter rollout" OR subsidies OR loans OR "voluntary programmes" OR "information campaigns" OR "research and development")
"Demand-side response" AND programmes AND policies
"Demand response" AND programmes AND policies
("Demand response" AND programmes AND policies) AND (tariffs AND interruptible/curtailment OR "dynamic peak pricing" OR "critical peak pricing" OR off-peak OR "demand bidding" OR time-of-use OR real-time)
"Energy efficiency" AND programmes AND policies
("Energy efficiency" AND programmes AND policies) AND (labelling OR “performance standards" OR "utility obligations" OR "supplier obligations" OR "market transformation" OR "smart meter rollout" OR subsidies OR loans OR "voluntary programmes" OR "information campaigns" OR "research and development")

Table 5: the process for creating the final search term

All of the search terms were used with and without Boolean logic to assess the impact of using the following symbols:

- “” (Parentheses): groups words or phrases within a statement to exclude others (e.g. “demand-side management”) (Hart, 2002, p. 153)
- AND: ensures that terms on both sides of this operator are present in the document (Hart, 2002, p. 153)
- OR: ensures that one of the terms on each side of this operator are present in the document (Hart, 2002, p. 153)

Boolean logic is the “system of combining words into a statement for searching” (Hart, 2002, p. 153). It is used to improve the search and to reduce the number of initial hits by including only those documents that include certain terms or phrases, rather than searching for individual words in the search string separately (as well as together). For example, the search string ‘(“demand-side management” AND policies) AND (labelling OR “performance standards”)’ ensures that only documents referring to DSM policy and labelling or performance standards are retrieved.

The ten search terms were developed based on the two broadest types of DSM: demand response and energy efficiency, as well as the overall term, demand-side management. Thus, for each of the three terms, the same search string was used – i.e. AND programmes AND policies, as well as the longer search strings that encompassed all of the relevant policies. The policies related to demand response were included in the demand response search string and the policies related to energy efficiency were included in the energy efficiency search string.

The * operator (Wildcard) can be used to ensure that the search includes the root of words (Hart, 2002, p. 153) but allows any variation in the end of the word to be included in the retrieved documents. For example, ‘policies’ would just search for the plural of ‘policy’ whereas ‘polic*’ would search for ‘policies’ and ‘policy’. However, it was discovered that this operator increases the number of irrelevant hits as well as the total number of initial hits, as it includes words other than those related to policy that start with ‘polic’, such as ‘police’ or 'policing'. Therefore, this operator was removed from the search strategy. Similarly, it was found that all 33 of the databases did not differentiate between 'policies' and 'policy'; hence, the use of the plural also retrieved documents containing the singular. This helped to reduce the number of search terms from ten to eight.

Eight search terms in 33 databases was calculated to take much longer than the one-year timeframe for data collection. Further refinement was undertaken to improve the search terms. It was identified that the term, 'demand-side management', was much more common than 'demand management' and the latter term was primarily used in Australia. However, an exploration of using both search terms in the Australian government and regulator websites indicated that using 'demand-side management' instead of 'demand management' had minimal impact on the results. The same principle was applied to 'demand response', which was much more commonly used in the literature than 'demand-side response'. The latter is increasingly becoming common in the European Commission and UK literature, though it is still not on a comparable level with the former to warrant a separate search term. An exploration of using both search terms in the UK government and regulator websites, as well as the European Commission Department of Energy website, indicated that using 'demand response' instead of 'demand-side response' had minimal impact on the results. The number of search terms was thus further reduced from eight to seven.

Comparing the results from the three remaining broad groups of search terms: demand-side management, demand response and energy efficiency both without policies listed and with policies listed, highlighted that similar results were being retrieved from each of the databases, and that it also appeared unnecessary to include three broad groups. Thus, it was decided to remove the two broad types of DSM and stick with the overall term, demand-side management. The potential loss of some relevant documents from not specifying energy efficiency or demand response in the search string was judged to be justified for reducing the scope of the systematic review to what was feasible in the one-year timeframe, but whilst remaining comprehensive. This further reduced the number of search terms from seven to three.

The two remaining long search strings (containing the demand response policies in one search string and the energy efficiency policies in the other search string) were deemed to be useful only if the policies were searched for individually alongside '"demand-side management" AND programmes AND policies' (i.e. "demand-side management" AND programmes AND policies AND labelling), rather than contained within one search string. In other words, including all of the relevant policies in one search stream had limited additional search benefits than just using "demand-side management" AND policies AND programmes. Although additional benefits may have resulted if the policies were searched for individually alongside the main stream, the total number of search terms would have then increased to thirteen (one for each of the twelve DSM policies included in the research and one without a specific policy included). For a future research project with greater resources, this would be the recommended approach for expanding the current systematic review.

Thus, the alternative approach was to reduce the number of search terms from three to one to: "demand-side management" AND programmes AND policies. This search term consistently performed as the most comprehensive term out of the ten original search terms in all 33 of the databases, as it produced the lowest numbers of initial hits and the highest number of relevant hits. As justified above, and which was confirmed by the panel of academics and the systematic review expert, including a larger number of databases and reducing the number of search terms to one comprehensive search term that used Boolean logic, was a more appropriate approach for answering the research questions within the time and resource constraints of the PhD than increasing the number of search terms and reducing the number of databases. Further search terms can be added to the review in further research (discussed in chapter six), as the systematic review can be readily updated and replicated, as sub-section 3.4.7 shows.

Following the completion of the full systematic review data collection, it was found that 54.9% of the documents (56/102) included the phrase, 'demand-side management', at least once within each document. If the phrase, 'energy efficiency', is included in the internal search, 99.0% of the documents (101/102) in the final sample either included the phrase, 'demand-side management', or the phrase, ‘energy efficiency’, at least once within each document. The one document that did not include either of the phrases referred to demand response (the other broad category of DSM alongside energy efficiency, as discussed in chapter two). The finding that 99.0% of the documents (101/102) discussed either ‘demand-side management’ or ‘energy efficiency’ (or both) when energy efficiency was not included in the main search string highlights that the 33 databases were ‘intelligent’ enough to link demand-side management and energy efficiency as related concepts. This provides further justification for the use of one search term, as the search term included relevant hits for energy efficiency (and demand response) policy evaluations as well.

The primary reason why 45.1% of the documents (46/102) did not include the phrase, ‘demand-side management’, is due to the lack of an internal search engine within the IEPEC and ECEEE databases, which were the databases with the most numbers of documents in the final sample (21 and 14 documents respectively). As such, 35 relevant and high quality documents were retrieved without inputting the search term. This reduced the number of documents that did not include the phrase, ‘demand-side management’, from 46 to 11 (10.8% of the sample), which further indicates the robustness of the search term used. The remaining eleven documents can be explained as per the discussions in the previous paragraph, where the databases were ‘intelligent’ enough to link demand-side management and energy efficiency as related concepts.

Due to the length of time it takes to conduct a systematic review, it is common to perform a review update once all of the pre-defined databases have been synthesised. In this research, Web of Science is used to do this, as it covers >12,000 journals and >160,000 conference proceedings, including direct translations of non-English publications. As discussed above, it is more comprehensive than Scopus and easier to use (such as the display and filtering functions) than Google Scholar. Within Web of Science, only journal articles (primary research and reviews) and conference proceedings were searched using the database’s filtering functions. Web of Science and the MCDM expert interviews database (discussed in sub-section 3.4.8) brought the total number of databases included in the systematic review to 35.

For stage four of the review protocol, a set of inclusion and exclusion criteria are defined to reduce the boundaries of the systematic review to what is feasible with the resources available. These are summarised in table 6 overleaf.

Inclusion Criteria
Documents that answer the research questions
Documents that pass the study quality assessment scale
Documents that discuss government-stimulated policies and programmes
Documents that are written in English
Documents that are freely accessible and downloadable from the internet

Exclusion Criteria
Documents that look at DSM policy but not mechanisms and impacts
Documents that do not pass the study quality assessment scale
Documents that discuss utility-stimulated DSM programmes
Documents that discuss trials, pilots, and small-scale R&D programmes
Documents that model the future potential of DSM
Documents that discuss theoretical aspects of DSM policy
Documents that are not written in English
Documents from hand searching
Documents from referrals
Documents from bibliographies and ‘snow-balling’

Table 6: inclusion and exclusion criteria for documents in the systematic review

The criteria in table 6 were developed from the quick scoping review and from discussions with the systematic review expert and the panel of academics. As the table shows, only documents that are written in English, freely accessible and downloadable from the Internet, and focus on government-stimulated policies are included in the systematic review. Although there is the potential for some relevant documents to be missed, such as older evaluations from the 1970s-1990s (which may not be available electronically), evaluations written in other languages that have not been translated into English, and commercially-sensitive DSM programmes, such as voluntary agreements between industries and system operators that have not been publicly evaluated, these are common issues faced in other disciplines that use systematic reviews. The issue of non-English academic publications is to some extent overcome through the inclusion of Web of Science (which includes direct translations of non-English publications). However, the issue still remains for non-academic publications.

A specific criterion of relevance to this research is the focus on government policies rather than utility-stimulated programmes or small-scale trials. Much of the DSM research conducted since the energy crises of the 1970s has focussed on utility activities or research programmes rather than the evaluation of national (or sub-national) government policy. This forms part of the justification for the research, as discussed in chapters one and two. However, utility programmes that were implemented as a result of government policy are included.

The inclusion and exclusion criteria are designed to outline the boundaries of the protocol and the statements in table 6 are adapted from those in the social science disciplines that use systematic reviews (particularly the practical guidance given in Gough *et al*., 2012; Harden and Thomas, 2005; and Petticrew and Roberts, 2006). The total number of initial hits of 4,360 documents (across the 35 databases), of which 660 were deemed relevant, was judged to be a large enough sample size to account for most of the relevant evidence on DSM policy evaluation. However, it is acknowledged that, although a comprehensive coverage of databases was undertaken, some evidence may not be captured by using one search term and by applying the inclusion and exclusion criteria listed in table 6.

3.4.5 Systematic Review Protocol: Stages 5-6

For stage five of the review protocol, the study quality assessment is the most important part of determining what is included in the final review sample. In the disciplines that use systematic reviews, study quality assessment scales are utilised to ensure that only high quality evidence is captured. In the medical sciences, the Jadad scale is widely used and involves a series of ‘yes/no’ answers being given to questions on whether or not the study was randomised, blind and withdrawal rates were stated (Jadad, 1998), as shown overleaf.

1) Was the study described as randomised? (+1 point)
2) Was the study described as double blind? (+1 point)
3) Was there a description of withdrawals and dropouts? (+1 point)

Additional points are given in the following instances:

- The method of randomisation was described and it was appropriate (+1 point)
- The method of blinding was described and it was appropriate (+1 point)

However, points are deducted in the following instances:

- The method of randomisation was described but it was inappropriate (-1 point)
- The method of blinding was described but it was inappropriate (-1 point)

Thus, it is a five-point scale with all documents scoring three points or less being excluded from the final sample (Jadad, 1998). However, the scale is not without its criticisms. Berger (2006) and Clark *et al*. (1999) argue that the scale places too much emphasis on blinding, can show low consistency between researchers and is over-simplistic. Furthermore, the Cochrane Collaboration argues that the Jadad scale puts too much emphasis on research reporting rather than research conduct (2011). Gough *et al*. (2012) similarly highlight this point in relation to systematic reviews in the social sciences:

"Although how well a piece of research is reported may be a proxy for the quality of the research itself, there is a danger of missing a good piece of research due to poor quality reporting; or wrongly judging a polished report as being indicative of a well-conducted piece of research." (Gough *et al*., 2012, p. 157)

Nevertheless, it is time-consuming and practically challenging to find out details on how the research was conducted for every document that is subjected to study quality assessment. Thus, the Jadad scale (or variations of it) remains one of the more well-known and utilised scales in the medical sciences. This thesis acknowledges the criticisms of the Cochrane Collaboration but due to time and resource constraints it takes the assumption that the reporting of research is a reasonable proxy for the quality of the research in each document. In the social science disciplines that use systematic reviews, this assumption is more commonly applied due to the lack of RCTs in fields such as crime, justice and social welfare. Thus, as in the energy policy field, studies are usually conducted using different methods in different contexts, and this approach to study quality assessment is more appropriate.

Study quality assessment scales used in the social sciences look at broader aspects of evidence quality than the arguably narrow approach of the Jadad scale. For example, Pawson *et al*. (2003)'s TAPUPAS framework, which assesses the quality of knowledge in social care, has seven dimensions:

- *Transparency:* are the reasons for the study clear?
- *Accuracy:* is the study honestly based on relevant evidence?
- *Purposivity:* is the method used suitable for the aims of the study?
- *Utility:* does the study provide answers to the questions it set?
- *Propriety:* is the study legal and ethical?
- *Accessibility:* can you understand the study?
- *Specificity:* does the study meet the quality standards already used for this type of knowledge?

The TAPUPAS framework is more useful than the Jadad scale for looking at DSM policy, as few policy evaluations include RCTs or controlled experiments. As previously shown, Pawson also developed the realist synthesis type of systematic review (2002b), which is utilised in this research, and highlights that his work on developing systematic reviews for use in social care research can be adapted for use in energy policy research. DSM policies vary not only in the category of DSM studied, but the way the policies were implemented, evaluated and administered (by governments, utilities, institutions, etc.). As such, policies and programmes are usually too heterogeneous to be statistically aggregated using meta-analysis and it is often inappropriate as they are undertaken in particular national (or state) contexts, which highlights that what works in one place may not work in another place (Petticrew and Roberts, 2006). Hence, DSM studies rarely include statements of randomisation, blinding and withdrawal rates. Instead, exploring the seven dimensions in the TAPUPAS framework is arguably more appropriate.

Other study quality assessment scales used in the social sciences attempt to combine the best of the scales in the medical sciences and the social sciences. For example, the EPPI-Centre (Evidence for Policy and Practice Information and Co-ordinating Centre) at the UCL Institute of Education in the UK developed a detailed appraisal tool for a systematic review of behavioural interventions to prevent sexually transmitted infections among young people (published in Shepherd *et al*., 2010). The appraisal included the following questions, which are split into assessment items and critical appraisal items and each document was subjected to the study quality assessment:

Assessment item:

1) J.1 Was a selection bias avoided?
2) J.2 Was a bias due to loss to follow-up avoided?
3) J.3 Was a selective reporting bias avoided?
4) J.4 Is the study sound?

Critical appraisal item:

5) E.17 Were steps taken to increase rigour/minimise bias and error in the sampling for the process evaluation?
6) E.18 Were steps taken to increase rigour/minimise bias and error in the data collected for the process evaluation?
7) E.19 Were steps taken to increase rigour/minimise bias and error in the analysis of the process data?
8) E.20 Were the findings of the process evaluation grounded in/supported by the data?
9) E.21 Please rate the findings of the process evaluation in terms of their breadth and depth
10) E.22 To what extent does the process evaluation privilege the perspectives and experiences of young people?
11) E.23 Overall, what weight would you assign to this process evaluation in terms of the reliability of its findings?
12) E.24 What weight would you assign to this process evaluation in terms of the usefulness of its findings?

What the above study quality assessment scale highlights is that scales need to be adapted to the research field in question, as directly applying the Jadad scale, TAPUPAS framework or equivalent to the energy policy field may result in few evaluations reaching the final sample, despite good quality policy evaluations having been conducted. Although the EPPI-Centre's critical appraisal tool is detailed and rigorous, its main drawback is the large amount of time required to appraise each document that reaches stage four of the systematic review protocol. For a systematic review with a team of researchers and a number of years to conduct the review, this approach might be more appropriate. However, within the resources and timescales of the PhD, a scale that has a similar degree of detail to the TAPUPAS framework is suitable (i.e. 6-7 questions). As systematic reviews have had limited attention in the energy policy field, there is a need to develop a scale that can be used to critically appraise energy policy evaluations. From the review of the systematic review literature, the following scale was developed and utilised in this research, which was peer reviewed by systematic review experts and published in the journal, *Energy Efficiency* (in Warren, 2014b):

- *Implementation:* has the process for policy implementation been clearly explained?
- *Evaluation:* has the process for policy evaluation been clearly explained?
- *Peer review:* has the document been peer reviewed or independently verified?
- *Transparency:* are there statements of copyright, regulatory compliance, and possible conflicts of interest present?
- *Reliability:* does the author/publishing organisation have a track record in the area?
- *Clarity:* where percentages are given, are the totals given?

The Warren scale has six main questions that need to be answered. However, discussions with systematic review experts and the review of the systematic review literature highlighted the importance of weighting the questions. There are a number of reasons to do this, for example, to adapt the focus of the scale to its relevance in directly answering the research questions, to ensure that documents are not overly penalised for excluding 'surface', reporting features (such as statements of conflicts of interest or giving totals where percentages are cited) when details on how the research was conducted are given (such as details on programme implementation and evaluation). Thus, a further eight points are included in the scale bringing the total number of points that can be achieved to fourteen rather than six (and the minimum score is zero). The full scale is shown in figure 16 below.

- **4 points:** Has the process for policy implementation been clearly explained?
    - *1 point: Have details on the policy implementer been given?*
    - *1 point: Have details on how the policy was designed been given?*
    - *1 point: Have details on how the policy was implemented been given?*
    - *1 point: Have details on why the policy was implemented been given?*
- **4 points:** Has the process for policy evaluation been clearly explained?
    - *1 point: Have details on the policy evaluator been given?*
    - *1 point: Have details on how the policy was evaluated been given?*
    - *1 point: Have details on the policy impacts been given?*
    - *1 point: Have details on policy success been given?*
- **2 points:** Has the document been peer reviewed or independently verified?
    - *1 point: Has the document been peer reviewed by a reputable expert?*
    - *1 point: Has the document been peer reviewed by two or more reputable experts?*
- **2 points:** Are there statements of copyright, regulatory compliance, and possible conflicts of interest present?
    - *1 point: Does the document give statements of copyright, regulatory compliance, or possible conflicts of interest?*
    - *1 point: Does the document acknowledge resource contributions from people or institutions?*
- **1 point:** Does the author/publishing organisation have a track record in the area?
- **1 point:** Where percentages are given, are the totals given?

Figure 16: *Warren Scale* for assessing quality in energy policy evaluations

Documents must score at least half of the available points (i.e. 7/14 points) to be included in the final sample. It is common in the systematic review literature to have scales with thresholds of half or two-thirds of the total number of points. The scale is designed so that at least one point must discuss either the process for implementation or the process for evaluation, as these are the most highly weighted parts of the scale (only six points can be reached if all other parts of the scale are maximised). In addition to this, at least one point must come from another part of the scale – i.e. a paper scoring four points for implementation and four points for evaluation (giving a total of eight points) will still fail to be included in the final sample if it does not have at least one point from the remaining six points.

The scale is designed to ensure that good quality evidence is not excluded due to the absence of 'surface', reporting features, such as statements of conflicts of interest. Thus, if the two sub-questions under the 'transparency' question are excluded, the document can still be considered high quality. Nevertheless, the 'transparency' questions are important as they give the reviewer an indication of potential influences on the study in question, such as funding bodies, contributors and conflicts of interest.

Similarly, the absence of peer review does not automatically result in the document being deemed poor quality. Although in the final sample very few documents passed the scale without scoring any points under the 'peer review' question (as this is still a strong indicator in many systematic review study quality assessment scales), theoretically it is possible for documents to pass without peer review if they score well in other areas (such as some government evaluations).

The 'reliability' question requires an examination of the track record of the author or organisation in the area. Not scoring a point in this section does not infer that the author or organisation is unreliable (though they can be), but instead that the author or organisation is new to the field.

For the 'clarity' question, the absence of totals where percentages are cited was added to the scale due to the frequency within which it is discussed in the systematic review literature as a form of best practice for the reporting of research.

The two most highly weighted questions (the processes for implementation and evaluation) cover study quality in more depth than the other questions. Validity, reliability and measurement bias are key aspects of assessing study quality. Validity refers to the accurate representation of the concept that is supposed to be measured (Kellstedt and Whitten, 2013, p. 101); reliability refers to the extent to which the measure is repeatable or consistent (Kellstedt and Whitten, 2013, p. 99); and measurement bias refers to the systematic over-reporting or under-reporting of values for a variable (Kellstedt and Whitten, 2013, p. 100). Validity can be broken down into internal and external validity.

Internal validity refers to the extent to which the intervention design, conduct, analysis and presentation have minimised or avoided biased comparisons of the interventions being evaluated, and external validity refers to the extent to which it is possible to generalise the results of the intervention to other settings (Jadad, 1998; Petticrew and Roberts, 2006, p. 127). The proposed scale above determines the internal validity of studies by obtaining details on who implemented and evaluated the policy, how it was designed and implemented, and details on the methodological approach for evaluating the policy. In the case of the latter, some degree of judgement is required on the part of the reviewer to judge the appropriateness of the methodological approach of the study in question. The extent to which the evaluator is qualified to undertake the evaluation is taken into account through the ‘reliability’ question and instead the question on who evaluated the policy has greater links to transparency rather than qualification. External validity is determined through the analysis of policy mechanisms and contextual factors (outlined in sub-section 3.5.2).

In summary, Jadad (1998) lists a number of relevant statements that should be considered when assessing the quality of interventions:

- The relevance of the research question
- The internal validity of the intervention
- The external validity of the intervention
- The appropriateness of data analysis and presentation
- The ethical implications of the interventions evaluated

The five statements are directly or indirectly captured in the two most highly weighted questions (the process for implementation and the process for evaluation). The strength of Jadad (1998)'s assessment is the focus on the way the research is carried out, thus to some extent dealing with the Cochrane Collaboration's criticisms of the Jadad scale. However, the lack of direct inclusion of these five points in the Jadad scale leads to a situation where a document can pass the study quality assessment if it states that the study was blinded, randomised and withdrawal rates are given (i.e. only indicating that the research reporting is of a good quality rather than the research being carried out to a high standard).

The Cochrane Collaboration's criticisms of the Jadad scale stem from a discussion of error reporting. In addition to the five points listed above, critical appraisal should examine types one, two and three errors. The Cochrane Collaboration focuses on types one and two errors, though ignores the importance of type three errors. Type one errors are where evidence of efficacy is accepted when the programme does not work, type two errors are where evidence is discarded when the programme actually works, and type three errors are where the wrong theoretical and conceptual framework has been used (Petticrew and Roberts, 2006, p. 128; Schwatz and Carpenter, 1999). Types one, two and three errors are not examined in the research through the study quality assessment scale but separately (in the same way as Jadad (1998)'s assessment). Following stage five, but before stage six, the documents are examined for types one, two and three errors through the conclusions that are made in each document and whether or not these match up to the evidence presented. This produces the final sample for the systematic review. However, types one, two and three errors are difficult to determine in energy policy evaluations (due to the lack of standardised practices for conducting evaluations), and as such, it was challenging to assess some of the documents on these aspects of critical appraisal. Consequently, more weight was placed on the study quality assessment scale in determining the final systematic review sample.

A potential bias faced in all scientific disciplines is the non-reporting of negative results. Although some academic journals have been set up specifically to publish negative results in order to overcome this issue (e.g. the *Journal of Negative Results*, the *All Results Journals* and the *Journal of Negative Results* *in Biomedicine*), most disciplines suffer from a publication bias (Peplow, 2014). In the DSM policy field, publication bias is not just an issue for academic evaluations. It is possible that a government may not want to publish an evaluation of a policy that has failed, or in some instances, the lack of governmental post-policy evaluations in the DSM field might be due to governments not wishing to monitor progress in case the policy is not performing as well as originally anticipated. Governments tend to receive little praise for policies that work well (as they are expected to perform well), but receive much criticism if they do not perform (Ekins, 2015). Thus, the incentive structures for evaluation are misaligned for governments. As Kunseler (2007) states, science can be abused if it is only used when it provides supportive evidence for the preferred policy direction. Furthermore, some countries where English is not the main language may not wish to translate and publish evaluations of policies that have not performed well. Further research is needed in order to explore the degree to which such publication biases exist in the DSM policy (and wider energy policy) field.

Industrial evaluations may be influenced by the agendas of the bodies providing funding for the evaluation of particular DSM policies. Evaluations funded by governments or government organisations may wish the evaluators to focus on the positive impacts and give less attention to the negative impacts. Evaluations funded by lobbyists may wish to emphasise the positive impacts where they want a policy to continue or to be expanded, or they may want to focus on the negative impacts where they wish for a policy to be discontinued and scrapped. Leopold *et al*. (1971) argue that policy evaluations undertaken by third parties should be presented to policy-makers in the form of disaggregated policy impacts, so that policy-makers have the full responsibility for the synthesis and conclusions on policy decisions. This can help to draw a line between the evaluators and the policy-makers in order to prevent value judgements and political interference (Kunseler, 2007). In summary, this thesis acknowledges that evaluations of DSM policies that have not performed well are likely to be under-represented in the systematic review sample. However, this is a problem facing most scientific disciplines.

For stage six of the review protocol, the final sample is obtained following the study quality assessment. Data are extracted and stored systematically for each document in the final sample. An Excel spreadsheet template was used to ensure consistency in data extraction between documents. The spreadsheet template was then copied for each of the 35 databases. The first tab of each spreadsheet contained the results of the filtering process (initial hits, hits after title skimming, hits after abstract skimming and applying the inclusion/exclusion criteria, and hits after the study quality assessment) to obtain the final sample for the database in question, and then a separate tab for each document was created with the same layout for data extraction. The data extracted for each document are broken down into three main boxes: background information (qualitative and quantitative), policy impacts (quantitative), and policy mechanisms (qualitative):

Background (qualitative and quantitative):

- Author(s) and institution(s)
- Full reference
- DSM policy year(s)
- DSM policy category
- Location (country/state)
- DSM policy implementer
- Sector targeted (industrial, commercial, public, residential)
- Number of participants (where relevant)
- Policy description (objectives, methods used, key conclusions)

DSM policy impacts (quantitative):

- Overall energy savings
- Overall carbon savings
- Peak load reductions
- Policy costs and benefits to utilities
- Policy costs and benefits to government
- Policy costs and benefits to consumers

DSM policy mechanisms (qualitative):

- Policy choice details
- Policy design details
- Policy coverage details
- Policy implementation details
- Policy evaluation details
- Policy success details

The background details are obtained mainly for classification purposes, but they also give useful insights for answering research question one on DSM policy implementation. The details on DSM policy impacts contribute to understanding what DSM policies have been successful in particular contexts, thus contributing to answering research question two. The details on DSM policy mechanisms directly answer research question two. The DSM policy categories that are examined (as discussed in chapter two) are:

- IPBDR = Incentive payment-based demand response
- PBDR = Price-based demand response
- MT = Market transformations
- IR = Infrastructure rollouts
- UO = Utility obligations
- LB = Labelling
- PS = Performance standards
- L&S = Loans and subsidies
- UBM = Utility business models
- R&D = Research and development programmes
- IC = Information campaigns
- VP = Voluntary programmes

Where policies are implemented together as a policy package, these are also examined. The policy packages that were analysed in the systematic review are discussed in chapter four, as these were not pre-defined. As per figure 9 and table 2 in chapter two, the DSM policy category level is examined rather than the specific DSM policy level in order to reduce the boundaries of the research to a more feasible, but still useful, level. To reiterate the hierarchy of DSM policy levels used in this thesis, three examples are given below:

Policy type: regulatory
DSM policy category: performance standards
Specific DSM policy: building codes

Policy type: market-based
DSM policy category: price-based demand response
Specific DSM policy: critical peak pricing

Policy type: information-based
DSM policy category: information campaigns
Specific DSM policy: energy audits

In order to answer the research questions, the level of the analysis needed to span across different countries and continents as well as different DSM policies. Thus, the number of DSM policies could be reduced without reducing the coverage of different DSM policies, or the spatial coverage of different countries/states, by moving from the specific DSM policy level to the DSM policy category level. A larger research project with greater resources should undertake the systematic review at the specific DSM policy level in order to obtain greater analytical detail for the design and implementation of future DSM policies.

3.4.6 Systematic Review Protocol: Stages 7-8

For stage seven of the review protocol, the systematic review analyses the findings primarily through synthesis configuration, which is the process of building up separate elements such as ideas into a connected whole (Gough *et al*., 2012, p. 180). The main part of the configurative synthesis is to develop theories of how and why the policies and programmes work (Snilstveit *et al*., 2012; Gough *et al*., 2012) in order to answer research question two. As such, the process involves both qualitative and quantitative analysis. Section 3.5 discusses the analytical process in more depth.

Although not listed in Harden and Thomas (2005)'s list of systematic review stages, other authors, such as Petticrew and Roberts (2006) and Gough *et al*. (2012), highlight the importance of determining the methods of disseminating the review before it has begun. Thus, for stage eight of the review protocol, this research primarily fits the *Instrumental model* of informing and influencing the decisions of policy-makers (Gough *et al*., 2012, p. 229). The research aims to provide evidence for policy-makers around the world on how and why DSM policies succeed or fail. For the academic community, journal and conference papers have been published; for the policy community, summary reports and consultation responses have been produced and presentations given; for the industrial community, summary reports have been disseminated and presentations given; and for the public, two short documentaries have been produced and made publicly available (these can be accessed at: https://www.you tube.com/watch?v=uHMYIHItInY and https://www.youtube.com/watch?v=__uB RI-jdTk).

3.4.7 Pilot Tests and Data Collection

An important part of undertaking evidence reviews, such systematic reviews and rapid evidence assessments, is to pilot test the robustness and replicability of the review protocol. The premise is that the review should be able to be conducted independently by different researchers that come to the same results. This also identifies the ease at which the review can be updated and extended in the future. As a result, two pilot studies were conducted with two different researchers in two different databases, as shown overleaf.

- Pilot study one:
    - *Resource and Energy Economics* (REE) (academic journal database)
    - Main researcher and researcher two
    - Search term: Demand side management policies programmes
- Pilot study two:
    - *Energy* (academic journal database)
    - Main researcher and researcher three

- Search term: “Demand side management” AND policies AND programmes

The first pilot study identified that the search, final sample and data/information extraction were generally the same. The only difference occurred during the study quality assessment where researcher two passed two papers and the main researcher passed one paper. However, the same paper was passed in both assessments and the researchers gave similar study quality scores (10/14 and 9/14 respectively). In observing the second paper in more detail, it is clear that both researchers identified the same paper as being of high quality, but the main researcher excluded it for reasons of its focus on DSM trials rather than DSM policy. The key data and information that were extracted were the same. It is interesting to note that researcher two was not an expert in DSM or policy analysis (though an expert in the energy and buildings field), but similar results were still produced. This highlights the ease within which the review protocol could be followed.

The second pilot study aimed to improve upon the feedback from the first pilot study to enhance the robustness of the review protocol, and to ensure that other confounding variables or chance were not the cause of the similarities in the results from the first pilot study. A different database, an adapted (more advanced) search term and a different researcher were used. The search term was adapted to ensure that DSM trials were excluded and that the number of initial hits could be reduced. Searching for ‘demand-side management programmes policies’ will search for each word individually as well as together or together with some of the other words in the search string but not others. This increases the number of initial hits and the number of irrelevant hits as discussed previously. Instead, Boolean logic was used to improve the accuracy of the search term, which was detailed in sub-section 3.4.4. ‘“Demand-side management” AND policies AND programmes’ ensures that all retrieved documents use the phrase, ‘demand-side management’, and the use of ‘AND policies AND programmes’ ensures that references to DSM but not DSM policy specifically are removed. The use of ‘programmes’ in addition to ‘policies’ was included in order to capture documents that refer to government programmes and government policies synonymously.

As researcher three was a DSM expert though not a policy analysis expert, he/she was more familiar with the field than researcher two. Both researcher three and the main researcher identified the same relevant papers and the quality scores for those papers were generally the same. Slight differences between researchers on the specific scores of some papers is likely to occur due to some study quality questions in the assessment requiring some degree of judgement on the part of the researcher, which is normal in systematic reviews in the social sciences (Dockerty *et al*., 2009; McConnell, 2010). The most important aspect is that the same documents pass the study quality assessment to be included in the final sample. Similar data and information were extracted between the main researcher and researcher three and the use of Boolean logic in the search term narrowed the number of initial hits to skim through and increased the number of relevant documents. It is important to note that none of the researchers involved in the pilot studies were required to analyse the extracted data and information, as the analysis took place after the one-year period for the systematic review data collection had been completed.

In summary, the search term, “demand-side management” AND programmes AND policies” was used in the following 35 academic, industrial and policy databases and websites (the MCDM expert database is discussed in sub-section 3.4.8 and did not require the use of a search term):

Academic (9 databases: journals):

- Energy Efficiency
- The Electricity Journal
- Energy
- Energy Policy
- Energy Economics
- Energy and Buildings
- Resource and Energy Economics (REE)
- The Energy Journal
- Electric Power Systems Research (EPSR)

Industrial (11 databases: conference databases and websites of institutions and industry groups):

- International Energy Program Evaluation Conference (IEPEC)
- European Council for an Energy-Efficient Economy (ECEEE)
- American Council for an Energy-Efficient Economy (ACEEE)
- International Energy Agency (IEA) DSM Programme
- Google (PhD literature review)
- Association for the Conservation of Energy (ACE)
- Open Grey
- National Grid
- British Institute of Energy Economics (BIEE)
- International Association for Energy Economics (IAEE)
- International Partnership for Energy Efficiency Cooperation (IPEEC)

Policy (13 databases: national and state government websites and regulator websites):

- US Department of Energy (DoE)'s Energy Citations Database
- UK Department of Energy and Climate Change (DECC)
- US Energy Information Administration (EIA)
- US Department of Energy (DoE)
- UK Office of Gas and Electricity Markets (Ofgem)
- California Public Utilities Commission (CPUC)
- European Commission Department of Energy
- US Federal Energy Regulatory Commission (FERC)
- UK National Audit Office (NAO)
- UK Public Accounts Committee for the House of Commons (PAC)
- China National Development and Reform Commission (NDRC)
- Australia Energy Regulator (AER)
- Australia Department of Industry

Other (2 databases: review update and MCDM analysis):

- Review update database: Thomas Reuters Web of Science
- Expert interviews database: Multi-Criteria Decision-Making (MCDM) analysis

The systematic review data collection and synthesis was carried out for one year between April 2013 and April 2014 and the pilot studies were carried out in March 2013. The systematic review data analysis was undertaken following the data collection and is discussed in section 3.5 and chapters 4-6. Figure 17 overleaf summarises the percentage of the initial hits that reached each filtering stage for the 35 databases and figure 18 summarises the number of documents at each filtering stage for the 35 databases: initial hits, hits after title skimming, hits after abstract skimming and applying the inclusion/exclusion criteria, and hits after the study quality assessment (the final sample).

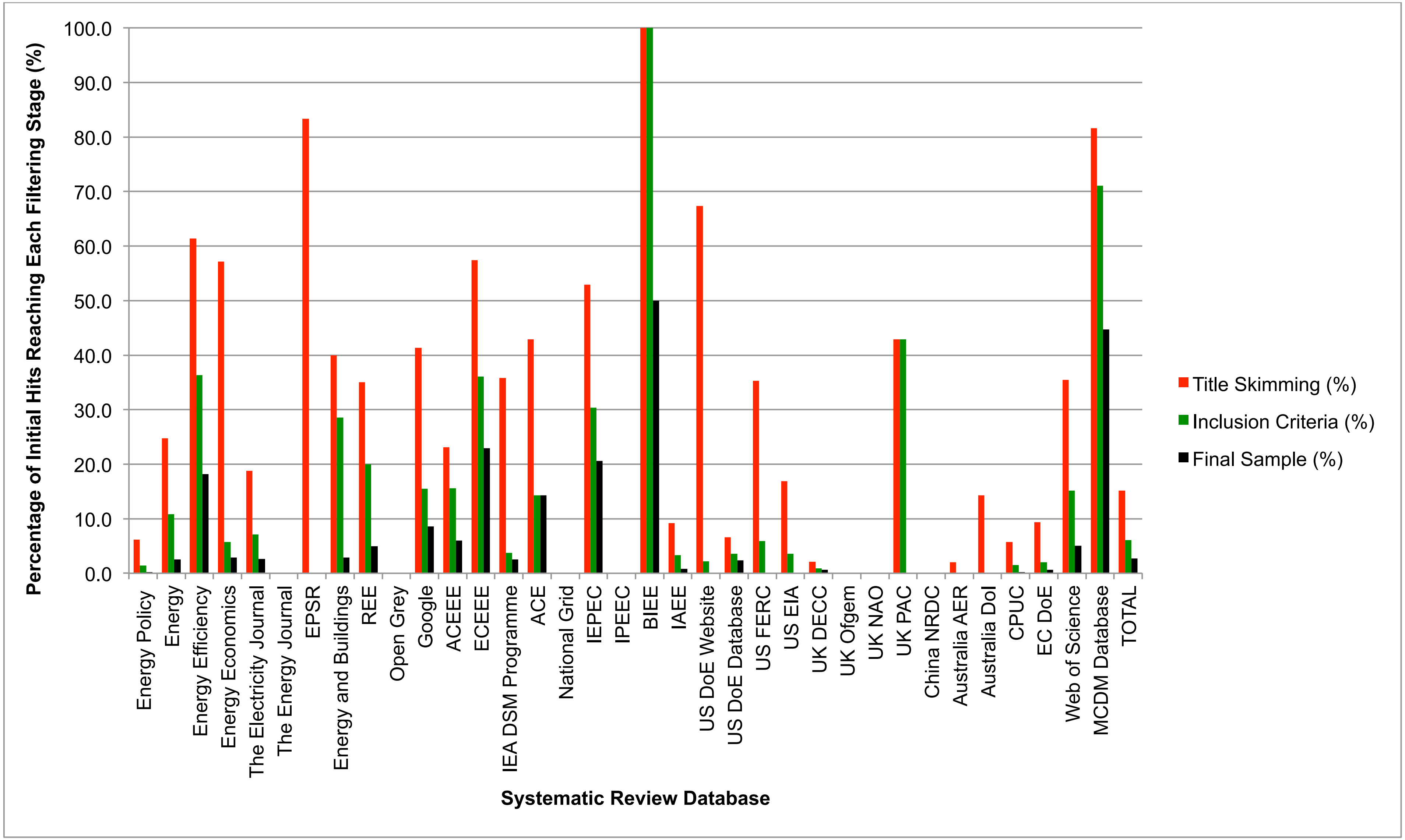


Figure 17: the percentage of initial hits reaching each filtering stage in the systematic review search

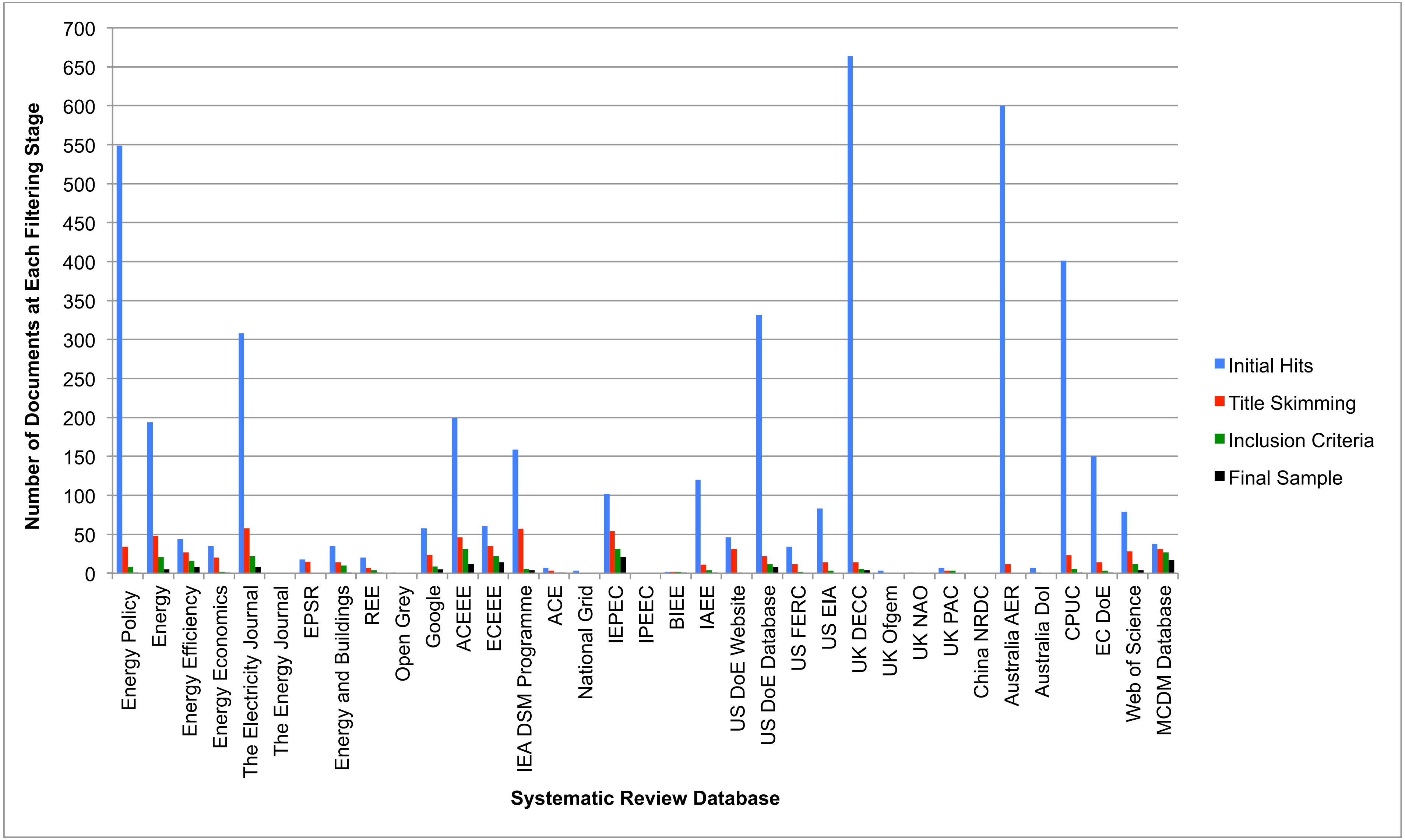


Figure 18: the number of documents reaching each filtering stage in the systematic review search

As previously mentioned, it is common to undertake a review update when carrying out systematic reviews. This is conducted at the end of the data collection period to ensure that new material published since the systematic review period began is captured. It also provides an opportunity to double check that relevant documents were not missed in the main period of data collection. As previously discussed, Web of Science was used as the main database for updating the review, as it covers >12,000 journals and >160,000 conference proceedings, including direct translations of non-English publications. However, as it does not cover industrial or policy databases, it was also necessary to undertake a review update in each of the 33 academic, industrial and policy databases included in the review. Web of Science was the last database to be synthesised and yielded four new and high quality documents. As the synthesis took place in April 2014, any new documents published since then are not included in the final analysis.

Figures 17 and 18 show that the databases that had the highest numbers of initial hits were not necessarily the same as those with the highest numbers of documents in the final sample (following the study quality assessment). Of the seven databases that had ≥8 documents in the final sample (i.e. the highest numbers of documents in the final sample), only three of these were also in the top seven databases by the number of initial hits: the *American Council for an Energy-Efficient Economy (ACEEE)*, *The Electricity Journal* and the *US Department of Energy (DoE)'s Energy Citations Database*. If the databases are examined by the percentage of the initial hits reaching the final sample, none of the top seven databases with the greatest numbers of initial hits are included in the top seven databases by the percentage of the initial hits reaching the final sample. The top seven databases in this respect have 8.6-50.0% of the initial hits reaching the final sample, whereas the databases with the highest numbers of initial hits have only 0.0-6.0% of the initial hits reaching the final sample.

If the discussion is extended to look at the percentage of the initial hits that are considered relevant (before the study quality assessment), similar results are produced. When ranked by the number of documents passing the inclusion criteria stage, only the ACEEE database and *The Electricity Journal* have both high numbers of initial hits and high numbers of documents that are considered relevant (i.e. in the top seven databases for the inclusion criteria stage). If the databases are then ranked by the percentage of the initial hits considered relevant, none of the databases in the top seven databases by the numbers of initial hits are in the top seven databases by the percentage of the initial hits considered relevant (having only 0.0-15.6% of documents considered relevant). For the top seven databases by percentage, 28.6-100.0% of the initial hits were considered relevant. This suggests that a high number of initial hits is a poor indicator of relevance. The top seven databases are examined (rather than the top five or the top ten databases) for consistency of discussion with the previous paragraph where seven databases had ≥8 documents in the final sample.

3/35 databases can be considered anomalies with respect to having no initial hits despite being identified as publishing prominently on DSM research: *The Energy Journal*, *International Partnership for Energy Efficiency Cooperation* (IPEEC) and China's *National Reform and Development Commission* (NRDC). All three of the databases were identified in the quick scoping review as publishing research on DSM rather than through referrals from the MCDM research participants. For *The Energy Journal*, it is clear that much research has been published on the technical, engineering and economic aspects of DSM, but not DSM policy. As discussed previously, the quick scoping review mapped out the databases publishing on DSM broadly rather than specifically on DSM policy (which was the focus of the systematic review). For IPEEC, the institution was mentioned in other documents in the quick scoping review that examined DSM (especially in non-academic documents). As such, the quick scoping review did not directly search within the IPEEC website but noted it as an important institution in the DSM field. For the NRDC, the department was discussed frequently in documents that looked at DSM in China (particularly in academic documents). As with IPEEC, the quick scoping review did not directly search within the NRDC website but considered it an important source of data and information due to its identification as the main department working at the national level on DSM in China.

In summary, the systematic review protocol was developed from a quick scoping review, from the development and testing of the search term(s), from feedback from a panel of five academics, from discussions with a systematic review expert, from two journal peer reviewers and from two pilot studies with two other researchers. The testing of the systematic review protocol suggests that it will have captured the majority of high quality DSM policy evaluations published in English (particularly in the last two decades), such that high confidence can be placed in the conclusions from the analysis of the documents, which are discussed in chapters 4-7.

3.4.8 Multi-Criteria Decision-Making Analysis

Multi-Criteria Decision-Making (MCDM) analysis was undertaken on a small-scale to act as a database to be fed into the systematic review. The triangulation of MCDM analysis into systematic reviews is a methodological improvement to mixed methods or qualitative systematic reviews, and has not yet been conducted in the disciplines that use systematic reviews. Methodological triangulation involves using more than one method to gather data (Denzin, 1978), in order to give “a more detailed and balanced picture of the situation” (Altrichter *et al*., 2008) and to cross-check information to produce more certainty in the results (O’Donoghue and Punch, 2003). MCDM analysis is gaining popularity in the energy policy field and is a quantitative method for undertaking face-to-face interviews. It aims to capture participant expertise and preferences for the ranking of various alternatives against how they perform on various criteria (UK CLG, 2009).

Traditionally, policy evaluation has focused on quantitative analyses, particularly cost-benefit and cost-effectiveness analyses, where a policy is successful if its benefits minus its costs are maximised (Nagel, 2002, p. 139). Nagel (2002) highlights that a policy is seen as a quantitative failure if more undesirable results are produced than desirable results, which are measured either by the intentions of the decision makers or by the objective effects regardless of intent (p. 139). Nagel (2002), a proponent of MCDM analysis for policy analysis, highlights alternative methods for analysing policy. He discusses mathematical optimising, econometric approaches, quasi-experimental approaches and behavioural process approaches, but argues that: “MCDM is closest to what good analysis and decision makers do implicitly” (p. 155).

There are over thirty different types of MCDM analysis (UK CLG, 2009). However, the general approach for conducting such analyses is highlighted in the stages below (taken from Nagel, 2002, p. 155):

1. List the available alternatives on the rows of a two-dimensional matrix
2. List the criteria for judging the alternatives on the columns of the matrix
3. Insert scores in the cells showing how each alternative relates to each criterion
4. Transform the scores to take into consideration that the goals may be measured on different dimensions
5. Aggregate the transformed scores across each alternative to obtain a summation score for each alternative
6. Conclude on which alternative or combination should be adopted

MCDM analysis is undertaken with identified experts in the field and seeks to rank the preferences of the experts as to which option should be chosen. In the case of this research, identified DSM policy experts would rank the twelve DSM policies discussed earlier in the chapter based on various performance criteria.

The main types of MCDM analysis that have been used in energy policy research to date are summarised in table 7 (adapted from UK CLG, 2009):

Multi-attribute models	Linear additive models	Analytic hierarchy process models	Outranking methods	Fuzzy sets models
Multi-attribute utility theory (MAUT)	Weighted sum method (WSM)	MACBETH	ELECTRE	VIKOR
Multi-attribute value theory (MAVT)	Weighted product method (WPM)	REMBRANDT	PROMETHEE	NSFDSS
		Analytic network process (ANP)		

Table 7: the main types of Multi-Criteria Decision-Making (MCDM) analysis methods

It is beyond the scope of the thesis to discuss each of the methods in detail, as the MCDM analysis is only a small part of the research. UK CLG (2009) provides a good overview of the different methods and how they can be utilised by governments and researchers, and thus is not reproduced here. The MCDM analysis is not analysed separately and is instead integrated as a database in the systematic review.

There were a number of reasons for conducting the MCDM analysis and to integrate it into the systematic review, rather than analysing the results separately. Firstly, it was a form of cross-validation for the systematic review results (though cross-validation was only based on the preliminary results as the MCDM interviews took place during the main period of data collection for the systematic review databases). It is important to note that the MCDM analysis would need to have a sample comparable to that of the systematic review sample of 102 documents (119 documents includes the MCDM analysis interviews) for the results to be representatively cross-validated. However, with a sample size of 17 (discussed below) this is not possible. Nevertheless, the interviews were also used to cross-validate the databases used in the systematic review to identify whether or not there were key databases missing. As discussed previously, an additional six databases were suggested by the research participants: Open Grey (database), National Grid (website), the US Department of Energy (DoE)'s Energy Citations Database (database), the US Federal Energy Regulatory Commission (FERC) (website), the UK National Audit Office (NAO) (website) and the UK Public Accounts Committee for the House of Commons (PAC) (website). However (as discussed previously), only the US DoE's Energy Citations Database produced relevant and high quality documents. Thus, the quick scoping review produced a more robust process for identifying relevant databases (and websites) for DSM rather than the referrals from the MCDM expert interviews.

In addition to cross-validation, the MCDM analysis aimed to synthesise and aggregate the results of the interviews into one expert database of 17 'documents' for direct integration within the systematic review, in order to qualitatively obtain perspectives on the key factors for DSM policy success and failure and to quantitatively obtain data on the success of different DSM policies.

The MCDM method used in this research is not one of the methods listed in table 7. Instead, it is an adapted, simplified and non-weighted version of the weighted sum method (WSM). Each expert was asked to rank (from 1-10) each of the twelve DSM policy categories based on eleven performance criteria. A score of ‘10’ is the highest value for performance and ‘1’ is the lowest value for performance. Scores for each policy are then summed to give an overall score. The overall scores of different policies are compared and then ranked (the highest scoring policy is ranked 1 and the lowest scoring policy is ranked 12 – where policies achieve the same overall score, the same ranking is given). Results from each expert interview are then aggregated to provide an overall ranking of policies. Thus, data from individual interviews are not analysed separately but included together as an aggregated whole.

In addition to the quantitative assessment, a qualitative element was added to the interview. As such, the MCDM interviews were both quantitative and qualitative, thus matching the mixed methods nature of the realist synthesis type of systematic review. The qualitative element aimed to answer research question two more directly, and involved three questions. The spreadsheet template (decision matrix) for the interviews is included in the Appendix (Appendix Figure 3). The DSM policy categories are re-produced below, alongside the quantitative and qualitative aspects of the MCDM interviews.

DSM Policy Categories:

- Incentive payment-based demand-side response (IPBDR)
- Price-based demand-side response (PBDR)
- Market transformations (MT)
- Infrastructure rollouts (IR)
- Utility obligations (UO)
- Labelling (LB)
- Performance standards (PS)
- Loans and subsidies (L&S)
- Utility business models (UBM)

- Research and development programmes (R&D)
- Information campaigns (IC)
- Voluntary programmes (VP)

DSM Policy Performance Criteria (Quantitative Assessment):

- Carbon emissions reduction
- Deferred investment in infrastructure
- Energy bill savings
- Government programme costs
- Overall energy savings
- Peak load reductions
- Utility programme costs
- Consumer active engagement
- Dealing with variable wind power
- Political ease of implementation
- Technology innovation and market development

DSM Policy Interview Questions (Qualitative Assessment):

1) From your experience, what have been the most successful policies?
2) How and why do you think these policies have been successful?
3) What are the key factors that may result in a DSM policy being unsuccessful?

Each interview equates to one document in the systematic review. 38 experts were identified from the literature review of 389 documents and were contacted by email. Of these, 17 experts agreed to participate (giving a participation rate of 44.7%): seven from industry, six from academia and four from policy. The participants signed a confidentiality form, which is included in the Appendix (Appendix Figure 2). The form gives an outline to the research project, how the meeting would be conducted and states that the participants' names would be kept anonymous but that they agree to be cited as an expert from their institution. The institutions are primarily based in the USA and the UK, which reflects the DSM policy evidence base, but also as a result of the resource constraints of the PhD to include other countries. Five interviews took place in the USA (New York and Washington DC) and twelve interviews took place in the UK (London, Oxford and Reading). The one-hour interviews took place face-to-face and were conducted between November 2013 and June 2014. A pilot test was conducted in November 2013 with another researcher to ensure that the interview process was robust and appropriate. In addition to this, the MCDM analysis approach was peer reviewed by a panel of five academics and two MCDM analysis experts. The institutions that took part in the research are listed below:

Research Participants

- Industry:
  - American Council for an Energy-Efficient Economy (ACEEE) (two interviews)
  - Edison Electric Institute (EEI) (one interview)
  - Association for the Conservation of Energy (ACE) (one interview)
  - Open Energi (one interview)
  - Scottish and Southern Energy (SSE) (one interview)
  - UK Demand Response Association (UK DRA) (one interview)
- Academia:
  - Imperial College London (one interview)
  - University of Oxford (four interviews)
  - University of Reading (one interview)
- Policy:
  - New York State Public Service Commission (NYSPSC) (one interview)
  - US Department of Energy (US DoE) (one interview)
  - UK Department of Energy and Climate Change (UK DECC) (one interview)
  - UK Office of Gas and Electricity Markets (UK Ofgem) (one interview)

The adapted MCDM method, which utilises a mixed methods approach to policy analysis in order to move away from solely focussing on policy impacts to policy mechanisms, triangulates well with review methods and it should be considered a useful technique for evaluating energy policies. Governments are usually time-constrained in terms of gathering evidence and evaluating policies and as such, a mixed methods rapid evidence assessment combined with a mixed methods MCDM analysis provides a practical and effective method for energy policy analysis when systematic reviews cannot be conducted due to time and resource constraints.

3.5 Analysis and Synthesis

3.5.1 Analysis Process: Research Question 1 – Policy Implementation

Research question one aims to answer the following question:

What DSM policies have been implemented around the world with high-quality documented evaluations?

Following the completion of the systematic review data collection (including the MCDM analysis and review update), the analysis for research question one is undertaken. The analysis process aims to look at the global spatial patterns of DSM policy implementation, but particularly in terms of the spatial distribution of high quality evaluations. Although spatial patterns are the focus of the research question, it is also interesting to briefly examine temporal patterns in how the number and distribution of evaluations has changed over time since the energy crises of the 1970s. The quantitative metrics that are required to answer research question one are listed below:

- Overall sample size in each of the four filtering stages during data collection
- Samples sizes per database in each of the four filtering stages
- Number and location of countries/states in the final sample
- Number of documents and evaluations per country/state
- Number of documents and evaluations per DSM policy
- Diversity of DSM policy implementation by country/state
- Frequency of different policy objectives for implementing DSM policies
- Number of documents and evaluations per country/state per decade since the 1970s

- Number and type of policy packages in the final sample

Each of the above metrics is discussed in detail in chapter four. It is important to note that one document does not equal one evaluation, and some documents contain multiple evaluations (either cross-country comparisons, cross-policy comparisons within the same country/state or cross-country and cross-policy comparisons). Chapter four discusses the analytical process and results for research question one in more detail.

3.5.2 Analysis Process: Research Question 2 – Policy Mechanisms

Research question two aims to answer the following question:

How and why do DSM policies succeed or fail, and what policies have been successful?

Following the completion of the analysis for research question one, the analysis for research question two is undertaken. This forms the central part of the thesis, which aims to identify the key success and failure factors for different categories of DSM policy and for different countries/states. The qualitative and quantitative metrics that are required to answer research question two are listed below:

- Defining policy success and failure
- Overall key success factors and failure factors
- Key success factors and failure factors by DSM policy: frequency
- Key success factors and failure factors by DSM policy: weighting
- Key success factors and failure factors by DSM policy: combined
- Key success factors and failure factors by country/state: frequency
- Key success factors and failure factors by country/state: weighting
- Key success factors and failure factors by country/state: combined
- Statistical associations between success factors
- Statistical associations between failure factors
- Overall success of different DSM policies

- Countries/states that have experienced success with each DSM policy
- Countries/states that have experienced failure with each DSM policy

Each of the above metrics is discussed in detail in chapter five. The main premise of the analysis process is to inductively identify the key success factors and the key failure factors for each DSM policy and country/state included in the systematic review sample. As the factors are not pre-defined, a first iteration of the systematic review analysis is undertaken to identify all of the factors across the 119 documents. A second iteration of the systematic review analysis is then conducted and the factors for each evaluation are categorised. The second iteration overcomes the issue of new factors identified in databases that are examined later on in the analysis process not being applied to earlier databases.

As discussed in chapter five, examining the frequency of discussion of different factors across the sample gives a good indication as to the importance of the factors. However, by itself it does not show how important a factor is in a given context (country/state and time period) within an evaluation. Thus, by looking at the weighting of different factors within each evaluation in the sample, their importance for particular policies and countries/states can be determined. Nevertheless, by itself, weighting cannot say whether or not the importance of a factor is replicated for the same policies in other countries/states. Hence, the weakness of weighting analysis is the strength of frequency analysis and vice versa. As such, there is strong justification for undertaking a combined analysis that filters out those factors that are not frequent and highly weighted. Despite this, the results of the individual frequency and weighting analyses are also presented in chapter five in order to provide more detailed analysis.

Pawson (2002b), who developed the realist synthesis approach to systematic reviewing, has been criticised for not giving practical guidelines as to how realist syntheses should be conducted and analysed (Dixon Woods *et al*., 2005; Warren, 2014b). As realist synthesis has had little practical implementation in the disciplines that use systematic reviews, there is limited literature on techniques for analysing data on programme mechanisms. This justifies the need to develop and describe new techniques in order to undertake realist synthesis analysis in the energy policy field. The scales and equations used to calculate frequency, weighting and the combined analysis are detailed and justified in chapter five.

The second part of research question two looks at what DSM policies have been successful (and unsuccessful) in the countries/states included in the systematic review. As shown in chapter five, a scale is developed based on the judgements of the evaluators in each document as to the overall success of the DSM policy under evaluation. As the documents are of a high quality and the evaluators are well qualified to undertake the evaluations, their judgements on overall policy success should be considered a reliable indicator. A scale of 1-5 is used where ‘1’ is a failed policy and ‘5’ is a highly successful policy. All policies scoring ‘4’ or ‘5’ are considered successful for the given country/state (as these are the highest scores). The overall policy success weighting for each DSM policy is then calculated by aggregating all of the individual policy success weightings given in each evaluation. In order to calculate an overall policy success score (across countries/states), the overall policy success weightings are combined with the results from the first part of the analysis (the frequency and weighting of success and failure factors). The equation used to calculate this is outlined and justified in chapter five. The final product from the second part of research question two is a list of successful DSM policies by country/state.

4 Chapter 4: DSM Policy Implementation

4.1 Data Collection

This chapter is concerned with answering research question one:

What DSM policies have been implemented around the world with high-quality documented evaluations?

Firstly, the chapter gives an overview of how the systematic review and MCDM (Multi-Criteria Decision-Making) analysis were conducted (sub-sections 4.1.1 and 4.1.2); secondly, it presents summary statistics on sample sizes and general patterns in the data (sub-section 4.2.1); thirdly, it discusses the spatial and temporal patterns of demand-side management (DSM) policy implementation (sub-sections 4.2.2 and 4.2.3); and fourthly, the conclusions for research question one are given. Discussions of implementation are based on the 119 high-quality documents included in the systematic review and not the 389 documents included in the literature review and the quick scoping review, which were undertaken prior to the systematic review. Hence, there are a number of governments that have implemented various DSM policies but which have not been documented in high-quality evaluations. Both the systematic review and the quick scoping review highlighted that even evaluations of lower quality (where limited resources are available) are seldom undertaken ex-post, after a policy has been implemented.

4.1.1 Overview: Systematic Review

The systematic review data collection was conducted over a one-year timeframe and included the synthesis of DSM evaluation documents from academic, industrial, and policy databases. The databases were selected based on being the main databases that have published research on DSM, determined through a quick scoping review undertaken prior to the systematic review. Following the filtering stages described in chapter three, a final selection of 98 documents were synthesised from 4,360 initial hits across 33 databases

(10 academic, 10 industrial and 13 policy). At the end of the one-year period for data collection (from 2013-2014), a review update was conducted using Web of Science (producing a further four high quality documents) and MCDM analysis was undertaken with seventeen key DSM policy experts (discussed in the next sub-section). This brought the total number of databases to 35 and the total number of ‘documents’ to 119 (where the MCDM analysis was included as a database with one expert interview equating to one document). It is important to note that one document does not equate to one DSM policy evaluation, as some documents included comparative evaluations of different DSM policies or different countries/states. From the 119 documents, 707 individual evaluations (690 written evaluations and 17 MCDM interviews) made up the total sample size. Of the 690 written evaluations, 45 evaluated policy packages and 645 evaluated individual policies. Comparative documents analysed either the same policy in different countries/states, evaluated different policies within the same countries/states or analysed different policies in different countries/states.

The analysis took place at the ‘policy category’ level (as per table 8 below) to ensure that the sample size for each policy was large enough to make adequate comparisons, as justified in chapter three. For example, instead of analysing how time-of-use tariffs performed in different countries/states, the broader category of price-based demand response was used as the level of the analysis. The list in table 8 overleaf summarises the total number of high-quality evaluations (out of 690 written evaluations) per DSM policy.

DSM Policy Category	Number of Policy Evaluations
Utility business models (UBM)	122
Information campaigns (IC)	118
Loans and subsidies (L&S)	100
Utility obligations (UO)	89
Performance standards (PS)	81
Incentive payment-based demand response (IPBDR)	62
Labelling (LB)	42
Price-based demand response (PBDR)	26
Market transformations (MT)	17
Research and development programmes (R&D)	17
Voluntary programmes (VP)	12
Infrastructure rollouts (IR)	4

Table 8: the number of policy evaluations in the sample by DSM policy category

Table 9 overleaf summarises the number of evaluations per policy at all of the DSM policy levels (policy type, DSM policy category and specific DSM policy). However, the main research analysis takes place at the DSM policy category level.

Policy Type	DSM Policy Category	Specific DSM Policy	Number of Policy Evaluations	Total by Policy Category
MARKET-BASED	Incentive payment-based demand response tariffs	Direct Load Control	8	**62**
		Interruptible/Curtailable Programmes	17	
		Demand Bidding/Ancillary Services Market	26	
		Emergency Demand Response	8	
		Relieving local network constraints	3	
	Price-based demand response tariffs	Time-of-Use Pricing	16	**26**
		Critical Peak Pricing	2	
		Real-time Pricing	7	
		Extreme Day Pricing	0	
		Extreme Day Critical Peak Pricing	0	
		Inverted Block Pricing	1	
	Market transformations	Energy efficiency barriers removal	17	**17**
REGULATORY	Infrastructure rollouts	Smart meter rollouts	4	**4**
	Utility obligations	Utility obligations	58	**89**
		White Certificate Trading	31	
	Labelling	Appliance labelling	14	**42**
		Equipment labelling	19	
		Building labelling	9	
	Performance standards	Appliance standards	18	**81**
		Equipment standards	29	
		Building standards/codes	34	

		Shareholder Incentives: Shared Benefits	2	
		Shareholder Incentives: Performance Targets	58	
		Shareholder Incentives: Rates of Return	2	
		Revenue Regulation	0	
		Revenue-Cost Recovery Mechanisms	3	
		Direct Incentives	0	
		System Benefits/Public Goods Charges	49	
	R&D programmes	Government-stimulated large trials	9	**17**
		Utility-stimulated large trials	0	
		Development and demonstration programmes	8	
INFORMATION	Information campaigns	General information programmes	36	**118**
		Energy audits	28	
		Information centres	16	
		Education/certification/technical support	29	
		On-bill information	1	
		Governing by example	8	
VOLUNTARY	Voluntary programmes	Industrial companies	8	**12**
		Power production, transmission, distribution	1	
		Commercial/public/institutional organisations	3	
5 Main Types	**12 Sub-Categories**	**44 Specific Policies**	**690**	**690**

Table 9: the breakdown of high-quality evidence by DSM policy

4.1.2 Overview: MCDM Analysis

MCDM analysis was conducted with seventeen key DSM policy experts. It is important to note that, as per the *Hierarchy of Evidence* concept discussed in chapter three, expert opinion is at opposite ends of the pyramid to systematic review evidence and in this respect it should not be given the same weighting in terms of evidence quality. However, five of the experts had written some of the documents included in the systematic review and the other experts had either contributed to the evaluations as co-authors or peer reviewers, or other authors in their institutions had written them. This increased the value of their oral evidence and also formed an important part of their identification as experts.

The purpose of the MCDM analysis was not to compare the results with that of the systematic review (for this type of analysis, the sample sizes would need to be comparable), but instead to form one of the 35 databases that make up the systematic review. The number of expert interviews is comparable to one of the larger written databases included in the sample. Both the systematic review and the MCDM analysis did not pre-define sample sizes, as this goes contrary to the process for conducting systematic reviews. In some methodologies that also take an inductive approach to science, saturation points are used. A saturation point is reached when no new or relevant information is obtained from further data collection (Saumure and Given, 2008). However, systematic reviews require the aggregation and synthesis of the majority of the relevant evidence on a given research topic (Warren, 2014b). Thus, the number of databases is pre-defined and the systematic review is not complete until relevant data have been extracted from all of the databases. With the MCDM analysis, the same process was applied to ensure synergies between both types of analysis. The total number of relevant experts identified from both the systematic review and the literature review was 38 (the population of interest). All 38 experts were contacted by email (as described in chapter three) and the overall response rate was 52.6% (20 experts responded). Of these, 44.7% (17 experts) agreed to participate (the sample). As such, the sample size was determined by the number of relevant experts that were willing to participate rather than continuing the number of interviews until a saturation point was reached.

Due to resource constraints, the majority of interviews took place in the UK (London, Reading, and Oxford) and the USA (Albany (New York State) and Washington DC). Although this may have resulted in a locational bias, this matches the systematic review evidence, where the USA and the UK have the greatest number of high-quality evaluation documents (as discussed below). MCDM analysis involves quantitative data collection of expert judgements during interviews. However, the method was adapted so that a qualitative element was included to explore the research questions further. The interviews lasted one hour and involved 30 minutes for ranking the twelve DSM policies by their performance on eleven criteria, and 30 minutes of semi-structured interview to answer the three questions shown below:

1. From your experience, what have been the most successful DSM policies?
2. How and why do you think these policies have been successful?
3. What are the key factors that may result in a DSM policy being unsuccessful?

Participants signed a confidentiality form agreeing that their names would be kept anonymous but that they would be referred to as an expert from their institution. The form is included in the Appendix (Appendix Figure 2). The institutions that took part in the research are listed below in the order that the interviews took place:

- New York State Public Service Commission (1 interview) – USA
- American Council for an Energy-Efficient Economy (2 interviews) – USA
- Department of Energy (1 interview) – USA
- Edison Electric Institute (1 interview) – USA
- Imperial College London (1 interview) – UK
- University of Oxford (4 interviews) – UK
- Department of Energy and Climate Change (1 interview) – UK
- Office of Gas and Electricity Markets (1 interview) – UK
- Association for the Conservation of Energy (1 interview) – UK
- University of Reading (1 interview) – UK
- Open Energi (1 interview) – UK

- Scottish and Southern Energy (1 interview) – UK
- UK Demand Response Association (1 interview) – UK

The interviews break down into six academic experts (from universities), seven industrial experts (from non-profit organisations and utility companies) and four policy experts (from national and state governments and regulators). The analysis does not make comparisons between academic, industrial and policy experts due to the small sample size. Instead, as previously discussed, the aim of the MCDM analysis was to be included as a database in the systematic review. Part of the justification for this was to capture knowledge that may have been missed in the systematic review and as a form of validation of the systematic review results. It is important to note that the three results chapters do not refer to the two analyses separately but together. Hence, in the thesis, ‘documents’ refers to the 102 written documents and 17 MCDM interviews, and the ‘systematic review’ refers to all 119 ‘documents’ following the triangulation of the systematic review and MCDM analysis data.

4.2 Global Implementation of DSM Policy

4.2.1 Key Statistics

The total number of countries and sub-national states included in the systematic review was 66, which covered six continents. This is broken down into 30 countries and 36 states (including regions within countries, provinces and states), which were determined inductively through the countries/states evaluated in the documents. The countries that had evaluations at a state-level (in addition to at a national level) were the USA, Canada, Belgium, China, India and Australia. The spatial patterns are discussed in sub-section 4.2.2. In the systematic review, no time limit was placed on the publication of the documents in the inclusion criteria. As such, documents spanned different decades, though the vast majority were conducted in the last ten years. The temporal patterns are discussed in sub-section 4.2.3.

Table 10 shows the breakdown of the documents included in the systematic review by database, with the number of documents per database (both in the final sample and in the initial hits) and how the databases compare when broken down into academic, industrial and policy databases. What is clear is that three databases (excluding the MCDM database) dominate: the *International Energy Program Evaluation Conference* (IEPEC) (21 documents), the *European Council for an Energy-Efficient Economy* (ECEEE) (14 documents), and the *American Council for an Energy-Efficient Economy* (ACEEE) (12 documents). All three are included as industrial databases, though it is important to note that all three institutions hold important global conferences where researchers from academia and government submit research in addition to those from industry. Academia refers primarily to universities, industry refers mainly to institutions, organisations and consultancies, and policy refers to governments (national and state-level) and regulators.

The evidence base is dominated by industry (59 documents), followed by academia (29 documents) and then policy (14 documents). An important finding here is that globally, few high quality evaluations of DSM policies have been conducted by governments, with the majority of evaluations being conducted by industry and academia.

Database Type	Database Name	Documents Included	Total Number of Hits
Academic Databases (Electronic) (10)	Energy Policy	1	549
Total Number of Documents:	Energy	5	194
25	Energy Efficiency	8	44
Total Number of Hits:	Energy Economics	1	35
5,035	The Electricity Journal	8	308
	The Energy Journal	0	0
	Electric Power Systems Research	0	18
	Energy and Buildings	1	35
	Resource and Energy Economics	1	20
	Open Grey	0	1
Industrial Databases (Electronic) (10)	Google (PhD Literature Review)	5	58
Total Number of Documents:	American Council for an Energy-Efficient Economy (ACEEE)	12	199
59	European Council for an Energy-Efficient Economy (ECEEE)	14	61
Total Number of Hits:	International Energy Agency's Demand-Side Management Programme	4	159
1143	Association for the Conservation of Energy (ACE)	1	7
	National Grid	0	3
	International Energy Program Evaluation Conference (IEPEC)	21	102
	International Partnership for Energy Efficiency Cooperation (IPEEC)	0	0
	British Association for Energy Economics (BIEE)	1	2

14	US Federal Energy Regulatory Commission (FERC)	0	34
Total Number of Hits:	US Energy Information Administration (EIA)	0	83
9,581	UK Department of Energy and Climate Change (DECC)	4	664
	UK Office of Gas and Electricity Markets (Ofgem)	0	3
	UK National Audit Office (NAO)	0	1
	UK Public Accounts Committee for the House of Commons (PAC)	0	7
	China National Development and Reform Commission (NDRC)	0	0
	Australia Energy Regulator (AER)	0	600
	Australia Department of Industry	0	7
	California Public Utilities Commission (CPUC)	1	401
	European Commission Department of Energy	1	150
Review Update (Electronic) (1)	Thomas Reuters Web of Science	4	79
MCDM Interviews Database (In-person) 1)	Multi-Criteria Decision-Making (MCDM) Analysis Interviews	17	38
Total Number of Documents	*Across 35 databases*	**119**	**4,360**

Table 10: the breakdown of documents included in the systematic review by database

4.2.2 Spatial Patterns

The evidence base is dominated by the USA, the UK and the US state of California. These countries and states are evaluated in 25, 22 and 20 documents respectively out of the 119 documents included in the systematic review. Other countries/states that perform well in this respect are France (13 documents), China (12 documents), Denmark (10 documents), Italy (10 documents), the US state of New York (9 documents) and Germany (8 documents). Figures 19 and 20 summarise the findings and split the results into countries and states, though the analysis for the three results chapters analyses the two sets together. It is important to reiterate that one document does not equal one country/state, as some documents included multiple evaluations of different countries/states and policies.

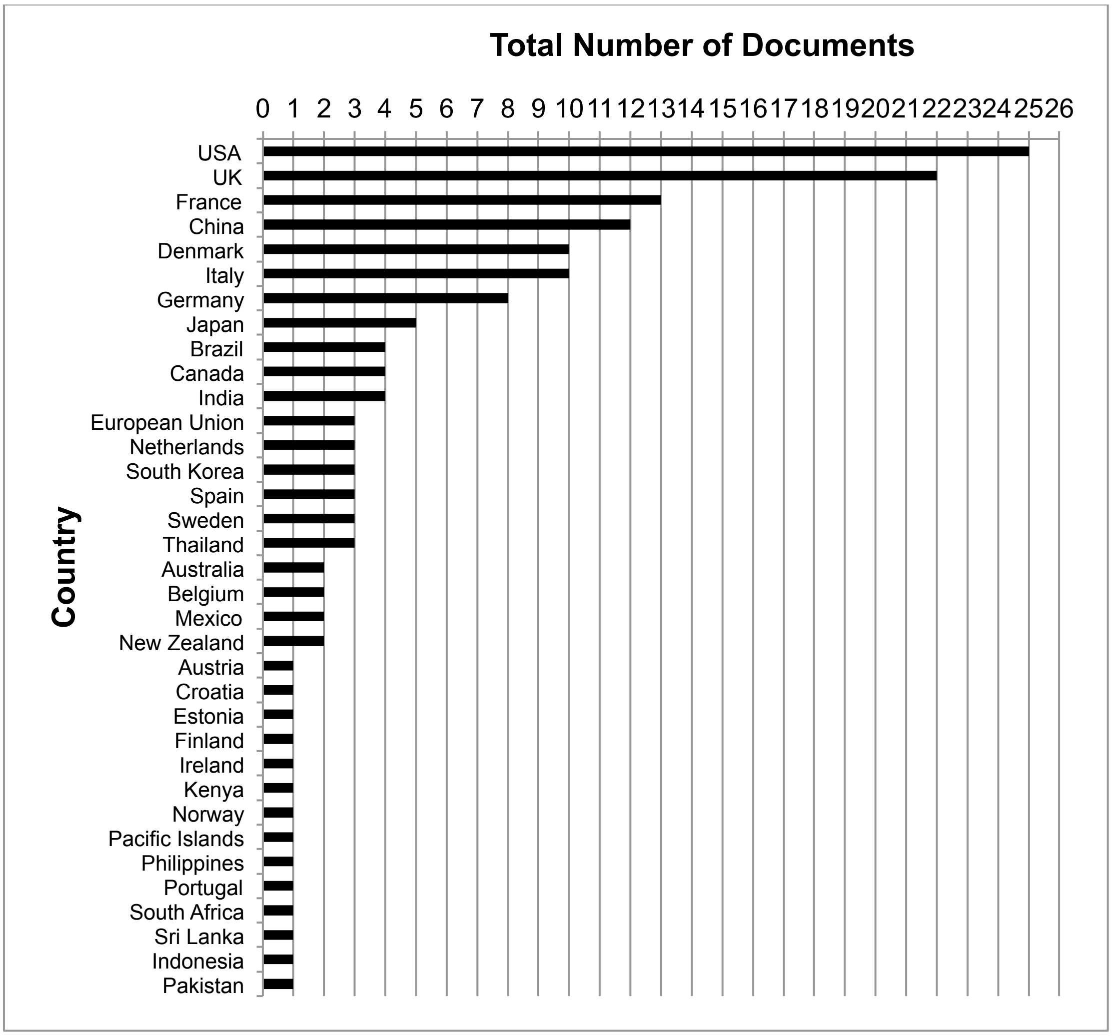


Figure 19: the evidence base of high-quality evaluations by country

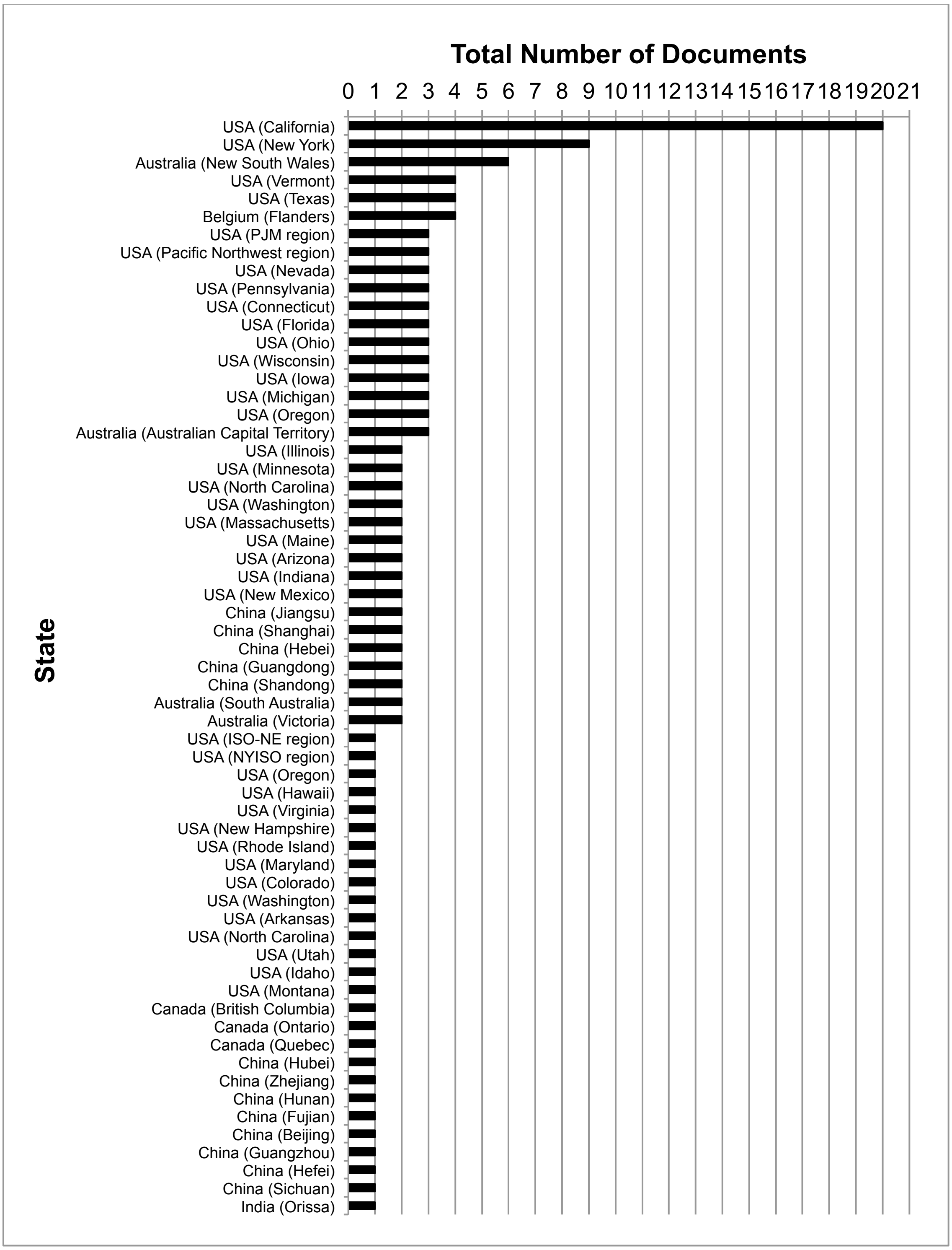


Figure 20: the evidence base of high-quality evaluations by (sub-national) state

The total number of countries in figure 19 is 35, but five countries do not have evaluations of sufficient depth for examination in chapter five, so are excluded from the main analysis (Austria, Finland, Portugal, the Pacific Islands and

Norway). As a result, 30 countries are included in the research. The total number of states in figure 20 is 59, but 23 states do not have evaluations of sufficient depth for examination in chapter five, so are excluded from the main analysis (the Canadian province of Quebec, the Chinese provinces of: Guangdong, Hubei, Hunan and Zhejiang, and the US states of: Arizona, Arkansas, Colorado, Hawaii, Idaho, Indiana, Iowa, Maryland, Minnesota, Montana, Nevada, New Mexico, North Carolina, Pennsylvania, Rhode Island, Utah, Virginia and Washington).

The top five countries in terms of the number of documents are the USA, the UK, France, China, and Denmark and Italy are joint-fifth. The top five states are California (USA), New York (USA), New South Wales (Australia), and Vermont (USA), Texas (USA) and Flanders (Belgium) are joint-fourth. All other states, regions and provinces have 1-4 documents. At a state-level, the USA leads overall with the number of documents far exceeding that of any other country at both a national or state-level. For example, 103 evaluations were conducted at a state-level in the USA, whereas 25 evaluations were conducted at a national level. Arguably, this implies that US state governments have been more proactive than the federal government, though much of the policy implementation and evaluation experience is located in the coastal states rather than the interior states. The total number of documents at a state-level by country is shown below:

- US states: 103 documents
- Chinese provinces: 18 documents
- Australian states: 13 documents
- Belgian regions: 4 documents
- Canadian provinces: 3 documents
- Indian states: 1 document

California dominates the evidence base at a state-level (with 20 documents) and although this is well known in the literature, this research has synthesised high-quality evidence to show that this is the case. Outside of the top five states, all other states, regions and provinces have 1-3 documents. As the analysis does not separate national and state governments, table 11 summarises the combined results for the top ten countries/states by the total number of documents (out of 119) in the sample.

Country/State	Total (Number of Documents)
USA	25
UK	22
California (USA)	20
France	13
China	12
Denmark	10
Italy	10
New York (USA)	9
Germany	8
New South Wales (Australia)	6

Table 11: the evidence base of high-quality evaluations of the top ten countries/states

DSM policies may be implemented individually or as part of a policy package. In the sample, nine policy packages were identified (i.e. had a frequency of one evaluation or more), which are given below with their abbreviations:

- **IPBDR/PBDR:**
  Policy package of *Incentive payment-based demand response/Price-based demand response*
- **UBM/MT:**
  Policy package of *Utility business models/Market transformations*
- **IC/L&S:**
  Policy package of *Information campaigns/Loans and subsidies*
- **PS/IC:**
  Policy package of *Performance standards/Information campaigns*
- **PS/LB/IC:**
  Policy package of *Performance standards/Labelling/Information campaigns*
- **PS/LB:**
  Policy package of *Performance standards/Labelling*
- **IC/L&S/MT:**

Policy package of *Information campaigns/Loans and subsidies/Market transformations*

- **PS/LB/UO/L&S:**
  Policy package of *Performance standards/Labelling/Utility obligations/Loans and subsidies*
- **VP/L&S:**
  Policy package of *Voluntary programmes/Loans and subsidies*

Despite the introduction of policy packages into the analysis, the evidence is still dominated by individual policy evaluations. Of the 690 written evaluations, only 45 included evaluations of policy packages. IC/L&S and PS/LB are the dominant policy packages with 13 and 10 documents respectively. Policy packages are examined in more detail in sub-section 4.2.4.

An interesting spatial comparison is to examine the number of different types of DSM policy that have been implemented by particular countries/states. The findings are summarised in figure 21. The figure highlights that the USA, California, the UK and China have the greatest diversity of implementation of DSM policies, which matches the findings for the total number of documents per country/state. The USA has implemented and evaluated fifteen different DSM policies, and the other three countries/states have each implemented and evaluated twelve different policies. The European Union (EU) (nine different policies) and Denmark (eight different policies) also perform well in this respect. The EU is categorised as a country in this research, despite being neither a country nor a state. However, it is included in the research, as it has implemented a number of important directives related to DSM, some of which are evaluated in the systematic review. The top four countries have implemented similar DSM policies though the USA, California and China have had greater experiences with policy packages than the UK.

Figure 21 shows that the implementation and evaluation of DSM policy has been limited globally with only 10/66 countries/states having implemented at least a third (7/21) of the individual DSM policies and packages under examination. Furthermore, the total number of countries in the analysis, where at least one high-quality evaluation has been conducted, is just 30 out of a possible 195 countries globally. Although some countries/states may have been missed due to the use of one search term, a quick scoping review of some of the countries that are not present, particularly those that have an important influence on global energy policy, such as Saudi Arabia and Russia, suggests that they have implemented DSM programmes, but they did not reach the final sample for the following reasons:

- The countries have not evaluated the implemented DSM policies
- Policy evaluations that have been conducted are not of a high quality
- Policy evaluations are not accessible (either not translated or not publicly available)
- DSM activities are utility-stimulated rather than government-stimulated

Figures 21 and 22 convey that the top four countries/states by the total number of evaluation documents are similar in having high-quality evaluations of many individual DSM policies, but show more diversity in the choice of policy packages. For example, seven individual DSM policies have been implemented by all four of the countries/states: incentive payment-based demand response, price-based demand response, utility business models, research and development programmes, information campaigns, performance standards, and loans and subsidies. In addition, two policies, market transformations and infrastructure rollouts, appear in three of the countries/states: the UK, the USA and California. Labelling also appears in three countries/states: China, the UK and the USA. The final two policies, utility obligations and voluntary programmes, appear in two countries/states each: the UK and the USA for utility obligations, and China and the USA for voluntary programmes. Thus, each individual DSM policy appears in at least two of the top four countries/states with most appearing in three or all four of the countries/states. However, out of the nine policy packages included in the research, no similar package is implemented in the top four countries/states. China has implemented IPBDR/PBDR, PS/LB and IC/L&S, the UK has implemented VP/L&S, and California has implemented UBM/MT, IC/L&S/MT and PS/LB/UO/L&S. The USA has not implemented any of the nine policy packages. Implementation in this respect refers to those countries/states that have implemented *and* produced high-quality evaluations of those policies.

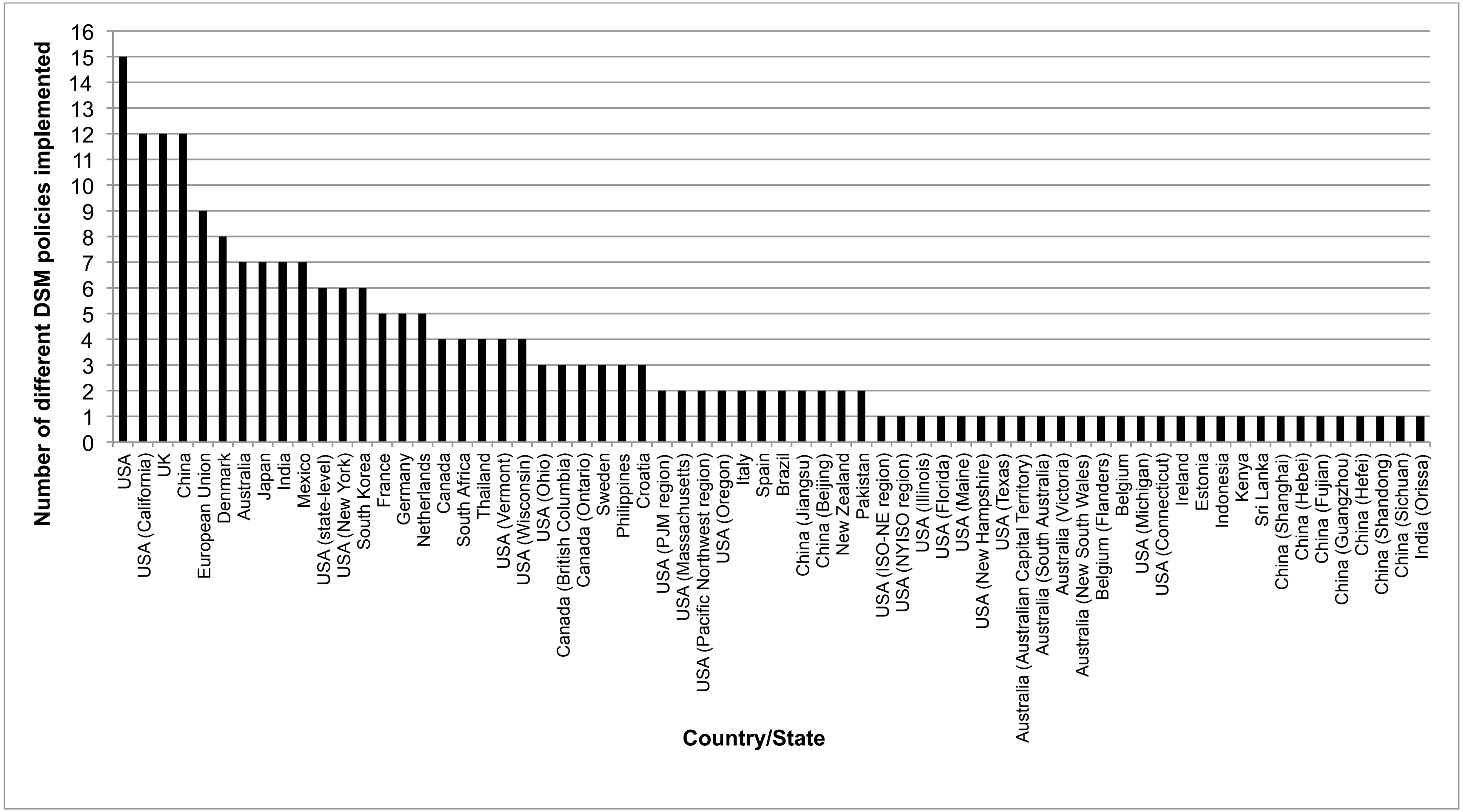


Figure 21: the number of different DSM policies implemented by country/state

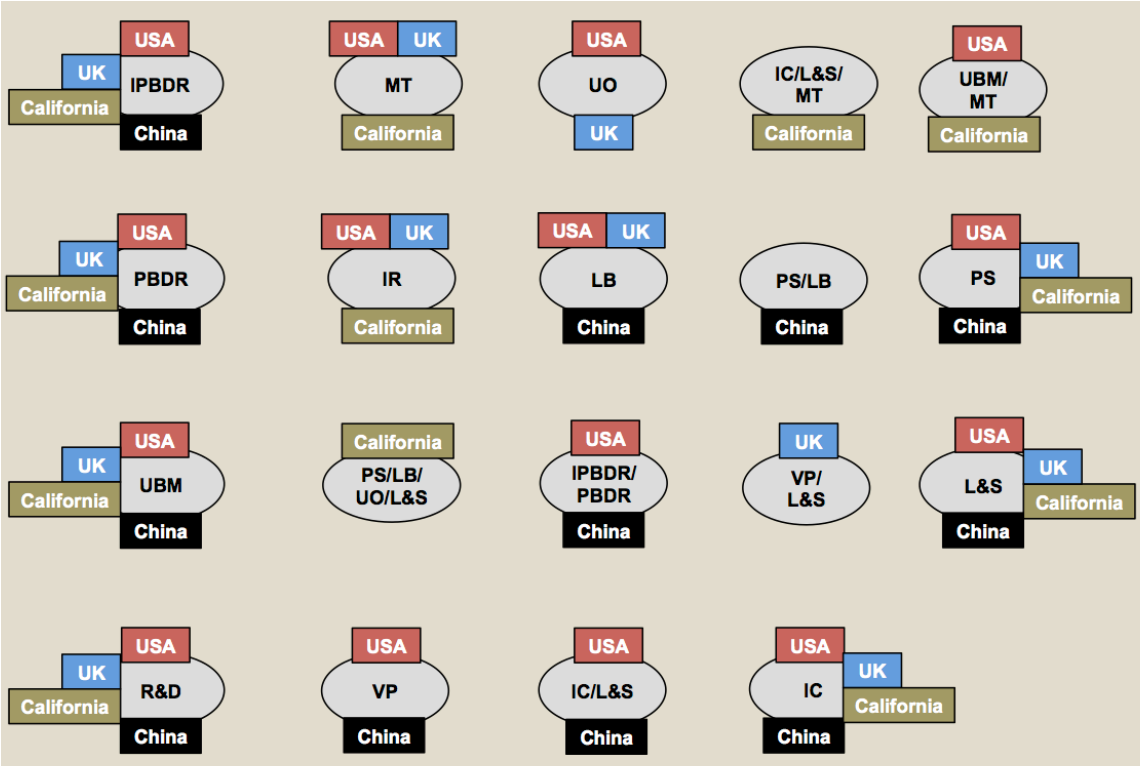


Figure 22: countries and states with the greatest diversity of DSM policy implementation

Figure 21 shows how the number of high-quality evaluations appears to be highest in the USA (at both national and state-level) and the UK, with Europe and east-Asia (particularly China) forming much of the remaining implementation and evaluation experience. Few evaluations appear for other continents, such as South America, Australasia (except state-level utility obligations in New South Wales), Africa and other parts of Asia. As the figure portrays, out of the 66 countries/states included in the systematic review, twenty-eight have implemented just one DSM policy, seventeen have implemented 2-3 DSM policies, eleven have implemented 4-6 DSM policies, six have implemented 7-9 DSM policies and four have implemented ≥10 DSM policies. If the top ten countries/states are examined rather than the top four (i.e. those that have implemented ≥7 DSM policies), the spatial diversity shows that the experience is dominated by North America (USA, California and Mexico), followed by Europe (UK, EU and Denmark) and east-Asia (China, Japan and India), with Australia representing the Southern Hemisphere. The bottom twenty-eight countries/states are primarily made up of states in the USA, Australia and China (twenty-two states and provinces). The six countries in this group are located in other parts of Asia, Europe and Africa (Indonesia, Sri Lanka, Estonia, Ireland, Belgium and Kenya).

DSM policies are implemented for a variety of reasons, such as to ensure energy security, to enhance economic productivity (through energy efficiency and new business opportunities), to reduce carbon emissions and to reduce consumer energy bills. The systematic review extracted the primary policy objectives for each policy evaluation (where stated) and examined their frequency of discussion across DSM policies and countries/states in the sample. Figure 23 presents the overall results.

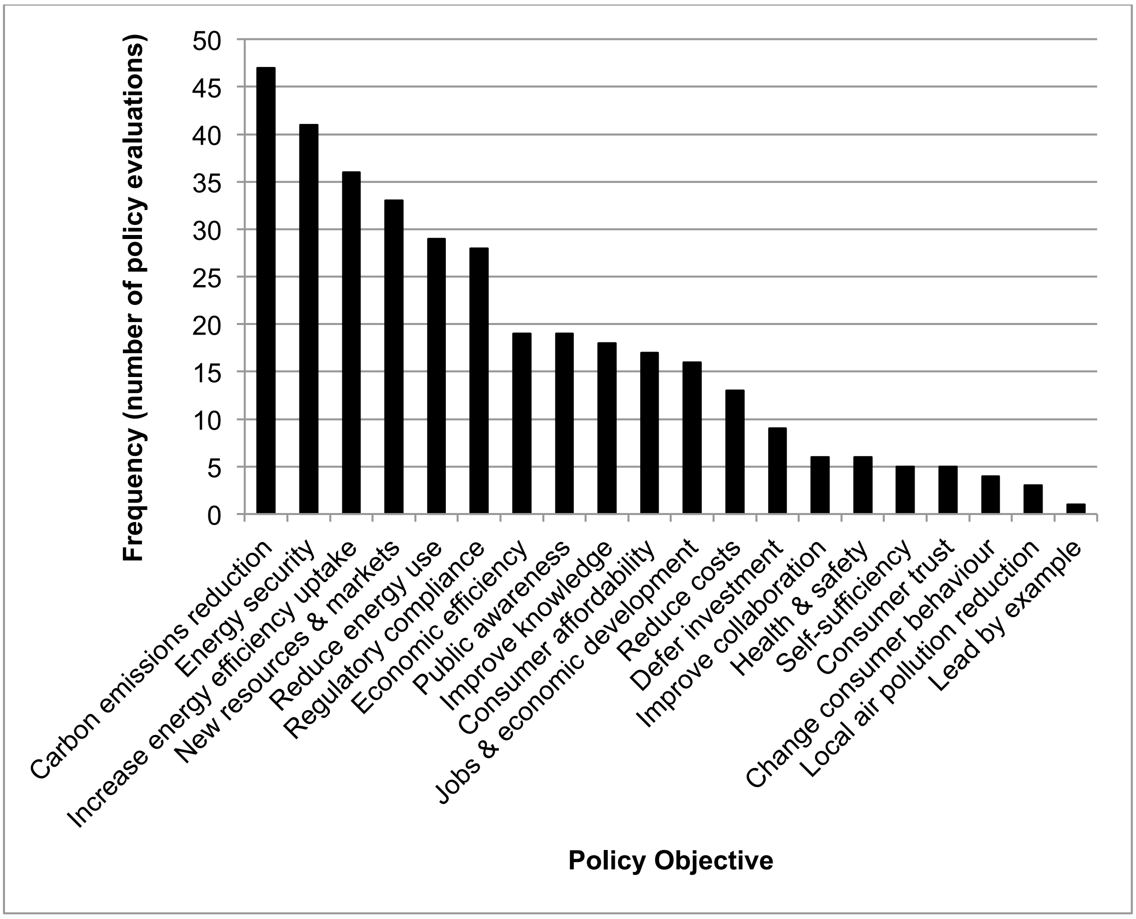


Figure 23: the primary reasons for DSM policy implementation

The policy objectives were determined inductively from the 102 written documents in the systematic review (as the MCDM interviews did not cover this). Where there were a number of evaluations within each document, policy objectives for each policy or country/state were noted (where they were stated). Figure 23 shows that overall, the top four reasons for countries/states to implement DSM policies are to reduce carbon emissions (47 evaluations), to ensure energy security (41 evaluations), to increase the uptake of energy efficient technologies (36 evaluations) and to create new resources and markets (33 evaluations). The first policy objective refers to all aspects of environmental improvement except for local air pollution reduction. However, all 47 evaluations referred to the reduction in carbon emissions either directly or indirectly. The second policy objective refers to the balance of supply and demand primarily at a national or state-level, though in one evaluation, it also referred to improving on-site energy security in the building. Peak load reduction was commonly mentioned in the 41 evaluations and is considered part of the ‘energy security’ policy objective rather than the ‘reduce energy use’ policy objective. This is due to the latter referring to an overall reduction in energy use rather than load shifting. The third policy objective refers to a range of specific goals, such as to increase the uptake of certain technologies by consumers, to improve the national/state building fabric and value through retrofitting, and to improve the efficient use of energy across the country/state. The fourth policy objective refers to the creation of new markets for DSM, to introduce DSM as a resource on an equal basis to supply-side options and to stimulate investment in DSM.

Although carbon emissions reduction and energy security appear to be the main policy objectives in most countries/states, there is some spatial diversity between continents. In North America, enhanced productivity through energy efficiency and new markets has been a strong driver in recent decades. For example, the *American Recovery and Reinvestment Act* (ARRA) of 2009 devoted significant financial support to DSM, which led to rapid job creation and retention as well as a reduction in energy use and greenhouse gas emissions (Doris *et al*., 2009). However, as discussed in chapters two and three, it is difficult to evaluate the impacts of national-level policies due to the challenges in establishing a reliable baseline of what would have happened in the absence of the policy. Enhanced productivity is captured through both the ‘economic efficiency’ and ‘new resources and markets’ policy objectives in figure 23.

In Europe, carbon emissions reduction and reducing consumer energy bills have become crucial drivers for DSM. For example, a number of the utility obligations in Europe (such as the UK, Italy, France and Denmark) are aimed at reducing carbon emissions by meeting energy savings targets and have sub-targets for low-income groups to contribute to reducing fuel poverty (Giraudet *et al*., 2012; Bundgaard *et al*., 2013a; Eyre *et al*., 2009). Fuel poverty refers to consumers living on a low income in a home that cannot be kept warm at a reasonable cost, as defined in the UK's *Warm Homes and Energy Conservation Act 2000* and reviewed in the Hills Fuel Poverty Review (2012). In Asia, particularly China, reducing local air pollution and preventing black-outs during crises have greatly expanded the development of DSM in recent years. The systematic review shows that much of the rapid growth in DSM has come from China (as shown in Bin and Jun, 2012; Pengcheng *et al*., 2012; and Zheng *et al*., 2012). The three documents that mentioned reducing local air pollution as a key policy objective were based on China. Preventing black-outs falls into the 'energy security' policy objective in figure 23 and was mentioned in the majority of the evaluations that focussed on China, as DSM is viewed as an important contributor to meeting the rapid growth in energy demand in the country.

The main conclusion from this part of the analysis is that the USA, California, the UK and China dominate the evidence base for DSM policy implementation and evaluation, both in terms of the number of evaluations and the number of different policies that have been implemented. Furthermore, the four most important reasons for implementing DSM policies overall are: carbon emissions reduction, energy security, increasing the uptake of energy efficient technologies, and developing new resources and markets. With respect to carbon emissions reduction, despite 66 countries/states having implemented DSM policies, Sorrell (2015) argues that larger and more rapid reductions in energy demand are required to meet this policy objective than have been achieved in the past.

It is beyond the scope of the research to examine how the policy objectives for DSM have changed over time for different countries/states. This is an area for further research, as discussed in chapter six. However, some temporal patterns of the evaluation evidence base are examined in sub-section 4.2.3.

4.2.3 Temporal Patterns

The analysis of spatial patterns showed that the DSM policy implementation and evaluation experience is dominated by the USA, the UK, California and China. Although not a crucial part of answering research question one, it is also interesting to look at temporal patterns in terms of the frequency of policy evaluations in different decades. As explained in chapter three, a key inclusion criterion of the systematic review protocol is that it only included documents that were accessible online. This potentially reduces the number of documents that could be included prior to the 1990s, as only those documents that have been digitised were accessible. This forms part of the reason why temporal analysis is not a central focus of this research, as well as due to resource constraints (time, funding and a research team to manually search libraries and government records). Furthermore, temporal analysis is not crucial for the exploration of the mechanisms behind the implementation, success and failure of DSM policies. Despite this, it is still interesting to make observations from the available data to see how the frequency of evaluations for the 66 countries/states included in the research has changed over time.

Figure 24 presents the results and categorises evaluations (contained within the documents) by decade based on their publication date and the period within which the policy was implemented. Where policy periods cross the threshold of decades, the decade with the most number of years for the policy period is used to prevent double counting (i.e. a policy that ran from 1979-1983 would be counted for the 1980s). In the case where a policy has an equal number of years in different decades, the publication date of the evaluation is used to categorise the decade.

The graph conveys that outside of the USA the evidence base is limited in the 1970s and 1980s. However, in the USA there was much DSM policy activity in this period, reflecting the energy crises of 1973 and 1979, which led to the *National Energy Conservation Policy Act* and the *Public Utility Regulatory Policy Act* (PURPA), which were introduced as part of the *National Energy Act 1978* (McNerney, 1998, p. 27) (as discussed in chapter one). In the 1990s, the global distribution of evaluations increases, particularly in Europe (notably Denmark, France, Germany, Italy, the Netherlands, Spain and Sweden) as the environmental agenda became more prominent, and east-Asia (notably China, India, Japan, South Korea and Thailand) as a result of energy security issues. The 2000s appears to be the greatest period for the number of evaluations undertaken with 42/66 countries/states showing more evaluations in this decade than in any other decade. In the 2010s, it is possible that the number of evaluations could overtake that of the 2000s if the analysis is extended post-2020, as less than half of the decade could be included in the systematic review at the time of data collection (2013-2014). If the prediction turns out to be correct, the results would show that there has generally been an increase in the number of high-quality evaluations over time. However, due to the digitisation bias discussed previously, further research is needed to validate this statement.

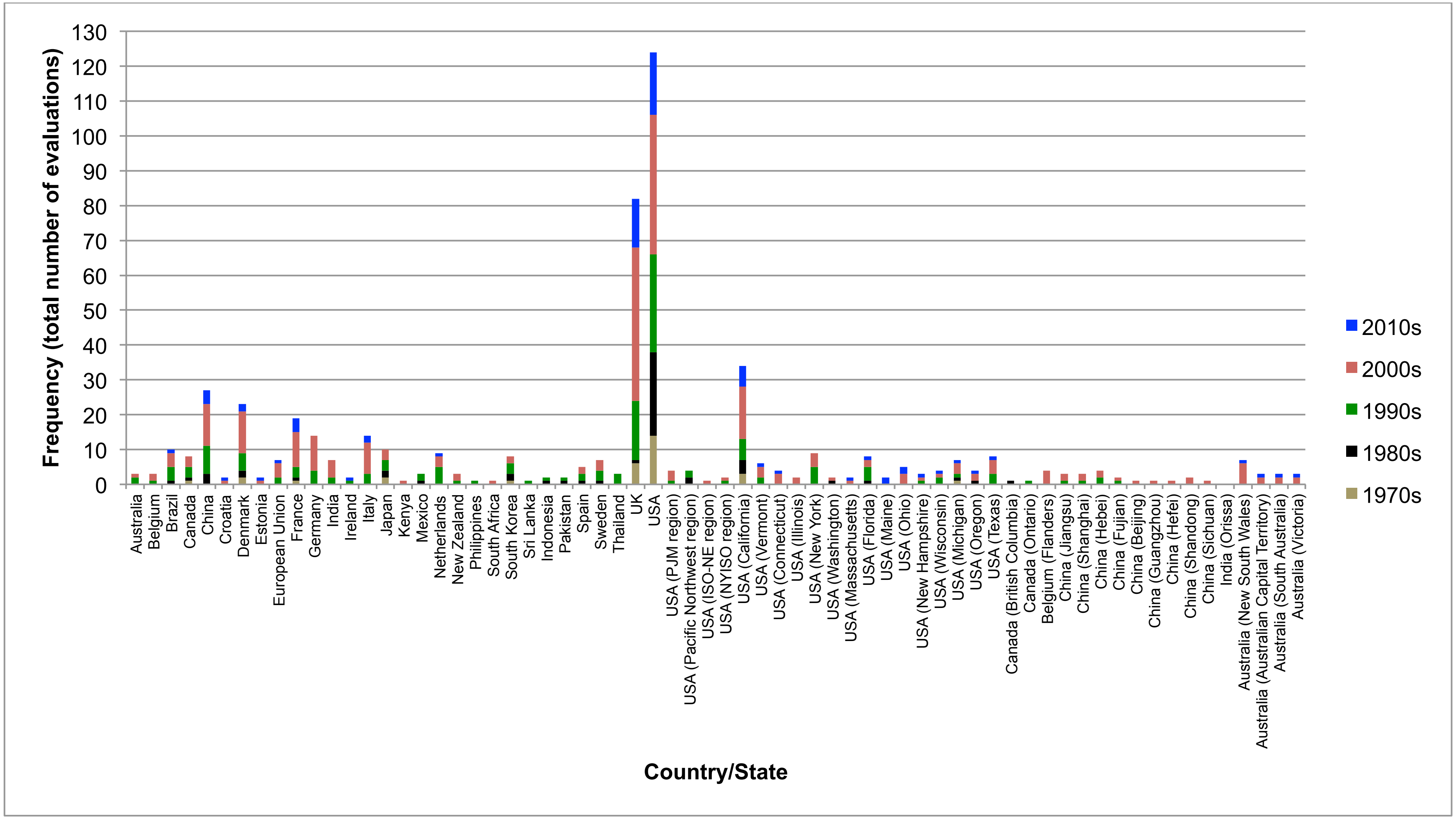


Figure 24: temporal analysis of DSM policy evaluations by country/state from 1970s to present

At a state-level, figure 24 portrays that the evidence base is dominated by US states across the decades. California has the greatest number of evaluations with a steady increase in frequency over time. Most of the evidence outside of California in the USA comes from the northeast (notably the New England and east coast states) and the northwest (notably the pacific north western states). For the northeast, a number of states begin evaluations in the 1990s and this increases in the 2000s. It is possible that the 2010s could continue the trend, as previously stated. In contrast, for the northwest, a number of states begin evaluations earlier in the 1980s and the frequency decreases over time. However, it is important to reiterate that the graphs only focus on the changing patterns of high-quality evaluations and not the actual implementation of DSM policies over time (the literature review and the quick scoping review conducted prior to the systematic review showed that DSM has generally increased across most regions over time).

In China, a number of provinces begin evaluations in the 1990s (notably Jiangsu, Shanghai, Hebei and Fujian) and this increases in the 2000s. The other provinces and regions included in the research begin evaluations in the 2000s (notably Beijing, Guangzhou, Hefei, Shandong and Sichuan), though it is too early to comment on whether or not the trend continues in the 2010s. In Australia, all four of the states included in the systematic review (New South Wales, the Australian Capital Territory, South Australia and Victoria) begin evaluations in the 2000s and this continues in the 2010s. In Canada, the two provinces included in the research only show evaluations in the 1980s (British Columbia) and 1990s (Ontario). In Europe, the only state to be included in the research is the Flanders region in Belgium, which begins evaluations in the 2000s though it is too early to comment on whether or not the trend continues in the 2010s (as per the other countries/states).

The main conclusion from this part of the analysis is that globally, the number of high-quality DSM policy evaluations appears to be increasing over time, particularly in recent decades (2000s and 2010s). This is especially the case in North America, Europe, and east-Asia.

4.2.4 Policy Clustering

The final part of the analysis for research question one takes the discussions of policy packages further to identify what DSM policies tend to be implemented together. Nine policy packages with one or more evaluations in the systematic review were identified, which are shown in figure 25. The figure highlights that the dominant policy package is when information campaigns (IC) are combined with loans and subsidies (L&S) (13 evaluations), followed by performance standards (PS) in combination with labelling (LB) (10 evaluations). In contrast, mixing these two combinations of policies together appears to be the least frequent policy package (performance standards in combination with labelling and information campaigns) with just one evaluation (in the Philippines).

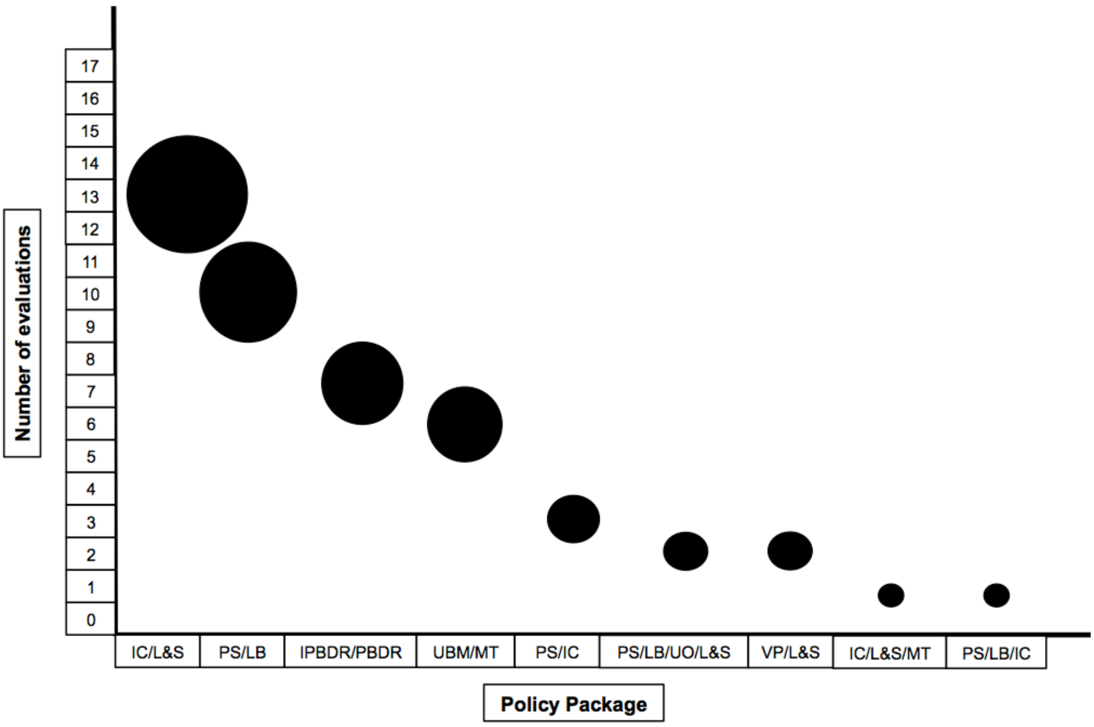


Figure 25: the number of evaluations by DSM policy package

The size of the bubbles in figure 25 simply reiterates the number of evaluations per policy package (the larger the bubble, the greater the number of evaluations). The IC/L&S policy package is popular due to the complementary nature of the policies. If a government offers loans or grants to consumers for

DSM measures (such as insulation or on-site generation), the policy is more likely to be successful if the incentives are well-marketed (policy success is explored in chapter five). Countries/states that have implemented this package include: China, Germany, Mexico, New Zealand, Sri Lanka, the USA, and the following US states: Illinois, Massachusetts, Maine, Ohio, New Hampshire and Wisconsin.

The PS/LB policy package is similarly popular due to the complementary nature of the policies. Introducing minimum energy efficiency standards for appliances, equipment or buildings are likely to be more successful in terms of manufacturing compliance and consumer education if they are clearly labelled (for example, with energy efficiency information, cost savings, carbon savings, etc.). The same argument applies to the PS/IC policy package where manufacturing compliance and consumer education should improve if performance standards are marketed clearly (though policy failure can still occur if other factors are not taken into account – policy failure is explored in chapter five). Countries/states that have implemented this package include: China, the EU, Pakistan, and the following Chinese provinces and regions: Shanghai, Beijing, Guangzhou, Hefei, Shandong, Sichuan and Jiangsu.

The IPBDR/PBDR policy package has the third highest number of evaluations (7) and reflects the pattern that countries/states that have implemented demand response tend to introduce both incentive payment-based demand response (IPBDR) and price-based demand response (PBDR) together, either individually or as a policy package. The former is concerned with providing direct payments or financial incentives to consumers to reduce demand during peak times, whereas the latter focuses on varying the price a consumer pays for energy at different times of the day or year (Albadi and El-Saadany, 2008). Much of the experience with demand response comes from the USA at a regional level through system operators, such as PJM, ISO-NE and NYISO. However, in recent years, China and some Chinese provinces have begun large-scale testing of demand response (notably Jiangsu, Beijing and Guangdong) (Wang *et al*., 2010).

The UBM/MT policy package is perhaps the most innovative of all the policy packages included in the systematic review. It is a longer term policy that aims to change the underlying business models of energy utilities (UBM) so that they can profit from demand-side options on an equal basis to supply-side options, while transforming the market for energy efficiency (MT). The package is arguably the most challenging to implement, particularly from a political point of view in liberalised energy markets, but it ensures that regulatory restructuring of utility incentives is complemented with changes in the market by overcoming market barriers. There has been little experience with this policy package outside of the USA, reflecting the limited practical implementation of changing utility business models, as discussed in chapter two. Around half of US states have implemented specific types of UBM (such as decoupling or system benefits charges) (NRDC, 2013), though the UBM/MT policy package has notably been implemented in the Pacific Northwest region (which covers Oregon and Washington), California, New York and Massachusetts.

Policy packages that involve voluntary programmes (VP) and loans and subsidies tend to be introduced for longer term objectives, such as to help stimulate the market development of energy efficiency products and services. The UK, and to a lesser extent India, have experience with this combination of policies. However, experience does not necessarily represent success, as explored in chapter five.

Finally, where countries/states implement more than two policies within a policy package (such as PS/LB/UO/L&S, IC/L&S/MT or PS/LB/IC), it often indicates a short-term response to an energy crisis, as previously highlighted in relation to the USA and California (such as the 2000-2001 electricity crisis in California).

The main conclusion from this part of the analysis is that globally, there are nine categories of DSM policy packages that have been implemented and evaluated, with the IC/L&S package forming the most frequent combination of policies, followed by the PS/LB package and the IPBDR/PBDR package. The IC/L&S package has the greatest spatial diversity, covering countries in east-Asia, North America, Europe and Australasia; the PS/LB package is dominated by experiences in Chinese provinces and European countries; and the

IPBDR/PBDR package has primarily been implemented at a regional level in the USA and to a lesser extent in China.

4.3 Research Question 1 Conclusions

The research question aimed to explore what DSM policies have been implemented around the world with high-quality documented evaluations. The systematic review identified twelve DSM policy categories and nine DSM policy packages. Of these, utility business models, information campaigns and loans and subsidies have the greatest number of evaluations in the sample with 122, 118 and 100 evaluations respectively. They are closely followed by utility obligations and performance standards with 89 and 81 evaluations respectively. At the other end of the scale, infrastructure rollouts has 4 evaluations and voluntary programmes has 12 evaluations. The number of evaluations by policy package is given below with their policy abbreviations:

DSM Policy Packages:

1. IC/L&S (13 evaluations)
2. PS/LB (10 evaluations)
3. IPBDR/PBDR (7 evaluations)
4. UBM/MT (6 evaluations)
5. PS/IC (3 evaluations)
6. PS/LB/UO/L&S (2 evaluations)
7. VP/L&S (2 evaluations)
8. IC/L&S/MT (1 evaluation)
9. PS/LB/IC (1 evaluation)

In total there were 690 policy evaluations (645 individual policy evaluations and 45 policy package evaluations) contained within 102 documents and 17 MCDM expert interviews. Thus, the final sample was 119 ‘documents’. The documents covered 66 countries/states (30 countries and 36 states). Of these, the USA, the US state of California, the UK and China have the largest evidence bases in terms of the number of policy evaluations and the number of different DSM policies implemented. Furthermore, since the energy crises of the 1970s it is clear that the USA (at a national and state-level) has dominated the evidence base over time. However, in recent decades (since the 1990s), Europe and east-Asia have begun to increase their experience with DSM policy implementation and evaluation. The 2000s appear to have been the greatest period for DSM policy evaluation in terms of the total number of evaluations conducted and the number of countries/states that have conducted evaluations. It is likely that the 2010s will continue to show an increasing trend of DSM policy implementation and evaluation over time, though the research needs to be extended post-2020 once policies for the decade have been implemented and evaluated.

It is clear from the results that policy packages only make up a small part of the evaluation evidence base (45 out of 690 evaluations). This could be explained by countries/states choosing to implement individual DSM policies, which usually entail lower administrative costs to governments, or by evaluation challenges. In the case of the latter, it is more complex to separate out the impacts for different parts of a policy package, and there is less experience in the DSM policy evaluation literature on the practicalities of evaluating policy packages. It is likely that both of these factors contribute to explaining the finding.

The success and failure of the twelve individual DSM policies discussed in this chapter are the focus of the next chapter, which forms the central part of the thesis. The chapter aims to answer research question two.

5 Chapter 5: Mechanisms for Success and Failure

This chapter is concerned with answering research question two:

How and why do DSM policies succeed or fail, and what policies have been successful?

Section 5.1 discusses how policy success and failure are defined in the thesis. Section 5.2 then describes and justifies the process for analysis in the chapter. The overall success and failure factors across DSM policies and countries/(sub-national) states are presented, and an analysis of the associations between factors is given. Section 5.3 examines the key success and failure factors by DSM policy and section 5.4 then undertakes the same process for each country/state. In each of these sections, worked examples are given to show the analytical process that was applied to all 12 of the policies and all 66 of the countries/states included in the research. Summary diagrams of the results for each policy and each country/state are provided. Section 5.5 describes and justifies the process for determining what countries/states have experienced success and failure with various DSM policies. Worked examples are provided to show the analytical process that was used and a summary of the results for each policy is presented. Finally, section 5.6 provides the conclusions to research question two.

5.1 Defining DSM Policy Success and Failure

5.1.1 Defining Policy Success

There is no single definition for determining the success of a policy (McConnell, 2010) and this is evident from the policy evaluation literature. For DSM, the research categorised four (non-exhaustive) means of identifying policy success, based on the literature review and the systematic review, as shown overleaf.

- Performance criteria (quantitative impacts)
- Stated success (qualitative judgement of evaluators)
- Stage in the policy process (success in design, implementation or post-policy evaluation)
- Underlying policy mechanisms (key factors for success and failure)

The analysis employed a combination of these categories in order to ensure a holistic definition was used. However, the primary focus was on the 'underlying policy mechanisms' and 'stated success' definitions, as these were more directly relevant to answering the research question. In determining the success of various DSM policies, performance criteria were also examined to ensure that the judgements of the evaluators matched the data that they presented (thus avoiding type one errors, as discussed in chapter three). Commonly used performance criteria in DSM policy impact assessments (as identified in the systematic review evaluations) are listed below:

Performance criteria:

- Carbon emissions reduction
- Deferred investment in infrastructure
- Energy bill savings
- Government programme costs
- Overall energy savings
- Peak load reductions
- Utility programme costs
- Consumer active engagement
- Dealing with variable wind power
- Political ease of implementation
- Technology innovation and market development

These performance criteria were used in the quantitative part of the MCDM analysis, as discussed in chapter three. Success based on the underlying policy mechanisms is generally qualitative and an area pioneered by Pawson (2002b), where the focus is not on impacts but on how and why an intervention succeeds or fails. The stated success of a policy refers to the qualitative overall judgement of the evaluator on whether or not the policy has succeeded or failed

(McConnell, 2010). This includes general statements on whether collected or estimated data empirically shows a policy to have been effective or not, discussions of whether or not the policy met its original overall objectives (McConnell, 2010), and statements on how the evaluators themselves define policy success.

Success based on the stage in the policy process refers to the evaluation of policies at the design, implementation or post-policy evaluation stage. It may include a combination of stages where the inputs and outputs of a policy are assessed. Examples include the degree of unintended consequences, whether or not outcomes meet objectives, the influence of one policy on another policy (and the degree of policy overlap), the cost-effectiveness of the policy, the degree of actual savings versus modelled savings, and the levels of free ridership and additionality.

5.1.2 Defining Policy Failure

In the thesis, the definition of 'policy failure' refers to policies not performing as well as originally anticipated. The definition is tied to the definition of 'policy success', as the primary focus is on the underlying mechanisms (failure factors) and the stated failure of policies by evaluators within evaluations. Similarly, performance criteria were examined to ensure that the judgements of the evaluators matched the data that they presented.

5.2 Success and Failure Factors

5.2.1 Analytical Process

The analysis presented in this chapter is based on three individual analyses. Firstly, the frequency of specific success and failure factors in the sample of 119 documents (representing 690 evaluations) is determined (frequency analysis). Secondly, the importance of those factors within each evaluation (as determined by the judgements of the evaluators) is identified (weighting analysis). Finally, the results of the frequency and weighting analyses are combined to establish the specific success and failure factors that are both frequent and highly weighted. Thus, the outputs from the combined analysis represent the main results. In this section, the analyses are conducted to identify the overall results, across DSM policies and countries/states. However, due to the nature of the calculations detailed below, the combined analysis requires information on a given policy or country/state under analysis, in addition to the information from the frequency and weighting analyses. As the analysis of the overall results is not related to a particular policy or country/state, the combined analysis is not conducted. However, for the rest of the chapter, the calculations for the combined analysis are undertaken in order to produce the final results by policy (section 5.2.5, averaged across countries/states) and by country/state (section 5.2.6, averaged across policies).

Frequency analysis was undertaken to look at the frequency of discussion of each success and failure factor in the sample. Where the discussions refer to 'documents', this represents the 102 documents and 17 MCDM interviews (119 'documents') included in the systematic review. Where the discussions refer to 'evaluations', this is the total number of individual policy or country/state evaluations included within the 119 'documents', which totals 690 evaluations (as discussed in chapter three). The key strength of frequency analysis is that it shows how widespread the finding is in the sample and whether or not factors identified for one policy implemented in a particular evaluation and context are present for the same policy in other evaluations and contexts. The frequency threshold shown in figure 26 was developed to differentiate factors that had a high frequency or a low frequency in the sample:

Factor Frequency Threshold:

1) **High Frequency:** ≥5 evaluations
2) **Low Frequency:** <5 evaluations

Figure 26: the *Factor Frequency Threshold* for determining the frequency of success or failure factors in policy evaluations

The level of '5' evaluations (not documents) as the threshold was determined inductively by examining the overall average frequency of discussion of each factor in the systematic review sample, which also required some degree of judgement (McConnell, 2010). The main limitation of relying on frequency analysis alone is that it does not identify how important the factors are for a given policy in a given context. Thus, weighting analysis was undertaken to overcome this issue. In order to calculate the weightings of success factors, a 1.0-3.0 weighting scale was used for each evaluation within each document. The scale is based on the qualitative emphasis that evaluators give to various success factors. To reduce the subjectivity of converting qualitative statements into quantitative data, specific words of emphasis were examined:

Factor Weighting Scale:

1) ***Score weighting 2.5-3.0 (Crucial):*** the following words are used in direct relation to the factor to strongly emphasis its importance: 'critical', 'crucial', 'very important', 'necessary', 'primary reason(s)', 'key', 'vital', 'central', 'essential', 'fundamental', 'decisive', 'significant' or equivalent
2) ***Score weighting 1.5-2.4 (Some Importance):*** the factor is included at the start of a list of factors and is frequently discussed though it is not strongly emphasised using any of the words for score weighting 2.5-3.0, or it is referred to using phrases such as: 'quite important', 'had some influence', 'played a role' or equivalent
3) ***Score weighting 1.0-1.4 (Small impact but not unimportant):*** the factor is included towards the middle or end of a list of factors without emphasis or discussion or it is indirectly inferred as a factor
4) ***No weighting (Unimportant):*** no weighting is given to the factor

Figure 27: the *Factor Weighting Scale* for determining the weighting of success or failure factors in policy evaluations

Weightings of 3 (high), 2 (medium), 1 (low) or 0 (unimportant) are assigned to each factor in each evaluation. However, when averages are calculated across policies and countries/states for each factor, figures to one decimal place are used for more detailed comparisons. One limitation of the technique is that evaluators may use language in different ways – for example, one evaluator's use of the word 'key' may be stronger or weaker than another evaluator's use of the same word. This is a challenge, but the literature is limited in this area and the proposed technique contributes to filling this methodological gap, as demonstrated in sub-sections 5.2.5 and 5.2.6.

The main limitation of just using weighting analysis is that it does not indicate how widespread the findings are in the sample. Instead, it identifies how important various success and failure factors are in specific contexts. Thus, the weakness of weighting analysis is overcome by undertaking frequency analysis and vice versa, and as such there is strong justification for combining the two analysis types to identify factors that are both frequent and highly weighted. In order to combine the two analysis techniques for use in the main results (sub-sections 5.2.5 and 5.2.6), the two-part equation shown in equation 1 below was developed:

Combined Frequency-Weighting Equation:

1) Frequency-Weighting combined analysis (FW_{pf}) = Policy Success weighting (PS_p) x (Policy Success Factor Frequency (PSF_{pf}) x Policy Success Factor Weighting (PSW_{pf})) / 10

2) Frequency-Weighting combined analysis percentage ($FW_{pf\%}$) = (Frequency-Weighting combined analysis (FW_{pf}) / Theoretical Maximum combined analysis (FW_{pfmax})) x 100%

In notation form:

1) $FW_{pf} = PS_p \times (PSF_{pf} \times PSW_{pf}) / 10$
2) $FW_{pf\%} = (FW_{pf} / FW_{pfmax}) \times 100\%$

Where $_{pf}$ is factor *f* for policy *p*.

Equation 1: the two-part *Combined Frequency-Weighting Equation* for determining the overall importance of success or failure factors in policy evaluations

The terms are explained below, which are followed by a worked example. The equations were undertaken for each of the success and failure factors discussed in sub-sections 5.2.2 and 5.2.3 for each DSM policy (sub-section 5.2.5) or country/state (sub-section 5.2.6).

Frequency-Weighting combined analysis (FW_{pf}):

FW_{pf} represents the values from combining the frequency and weighting analyses for a given success or failure factor for a given DSM policy (section 5.2.5) or country/state (section 5.2.6).

Policy Success weighting (PS_p):

PS_p represents the stated success of a given policy through the qualitative judgements of the evaluators as to the overall performance of the policy. To calculate PS_p for each policy, a scale of 1-5 is used (scales of 1-5 are widely used in the field, particularly in surveys, such as the commonly used five-part *Likert scale*, which was developed by Likert, 1932), as shown in figure 28 below:

Policy Success Weighting Scale:

1 = A failed policy that met none of its original objectives
2 = A poorly performing policy that met few of its original objectives
3 = An average performing policy that met most of its original objectives
4 = A policy that performed well and met all of its original objectives
5 = A highly successful policy that performed beyond its original objectives

Figure 28: the *Policy Success Weighting Scale* for determining the success of policies in policy evaluations

An average is then taken across the sample for each policy. The analytical process for PS_p is explained in more depth in section 5.3 in the discussion of successful DSM policies.

Policy Success Factor Frequency (PSF_{pf}):

PSF_{pf} represents the frequency of a given success or failure factor *f* for a given policy *p* in the sample of 690 evaluations, as determined in the frequency analysis.

Policy Success Factor Weighting (PSW_{pf}):

PSW_{pf} represents the importance of a given success or failure factor *f* for a given policy *p* in the sample of 690 evaluations, as determined in the weighting analysis.

Frequency-Weighting combined analysis percentage ($FW_{pf\%}$):

Like FW_{pf}, $FW_{pf\%}$ represents the values from combining the frequency and weighting analyses for a given success or failure factor for a given DSM policy (section 5.2.5) or country/state (section 5.2.6). However, it compares the result to the theoretically maximum result that could be achieved (see the explanation for FW_{pfmax} below) and writes the result as a percentage of this. The percentage is used as the final result for determining whether or not a given factor *f* is both frequent and highly weighted for a given policy *p*.

Theoretical Maximum combined analysis (FW_{pfmax}):

FW_{pfmax} represents the theoretically highest score that could be achieved for a given factor *f* for a given policy *p*. This is calculated by multiplying the frequency of discussion of a given policy *p* in the sample with the theoretically maximum possible success weighting of the policy (i.e. 5.0 as per the scale in figure 28), and then multiplying the resulting value with the overall success weighting of the policy in the sample. The explanation is visualised in equation two below:

Theoretical Maximum combined analysis Score (FW_{pfmax}) = Policy Success weighting (PS_p) x (Policy Frequency (P_p) x Theoretical Maximum Policy Success Weighting (PS_{pmax})) / 10

In notation form:

$$FW_{pfmax} = PS_p \times (P_p \times PS_{pmax}) / 10$$

Where P_p is the frequency of the policy in the sample of 690 evaluations and PS_{pmax} is the theoretical maximum policy success weighting of 5.0.

Equation 2: theoretical maximum combined analysis score

In part one of the combined frequency-weighting equation and the theoretical maximum combined analysis equation, dividing the resulting values by ten is undertaken in order to produce a more comparable and manageable scale for categorising success or failure factors. In part two of the combined frequency-weighting equation, the final value is multiplied by 100% in order to obtain a percentage of the theoretically maximum score that is achieved by a given success or failure factor. The level of ‘5.0%’ as the threshold for the combined analysis draws parallels to the frequency analysis threshold and was similarly determined inductively by examining the average combined analysis scores of the various success and failure factors, which also required some degree of judgement (McConnell, 2010). The scale shown in figure 29 below was developed to differentiate factors that are both frequent and highly weighted from those that are frequent but have a low weighting, are highly weighted but have a low frequency, or are infrequent and have a low weighting.

<u>*Factor Frequency-Weighting Combined Scale:*</u>

1) ≥10.0% of the theoretical maximum = Crucial factor
2) 5.0-9.9% of the theoretical maximum = Important factor
3) <5.0% of the theoretical maximum = Unimportant factor

Figure 29: the *Factor Frequency-Weighting Combined Scale* for determining the overall importance of success or failure factors in policy evaluations

A second level of importance was created in order to identify those factors that are ‘crucial’, in addition to those that are ‘important’. This also better aligns the scale with the three-part (1-3) *Factor Weighting Scale* shown in figure 27. If the factor achieves ≥10% of the theoretical maximum it is considered a ‘crucial’ factor, if the factor achieves 5.0-9.9% of the theoretical maximum it is considered an ‘important’ factor and if the factor achieves <5% of the theoretical maximum it is considered an ‘unimportant’ factor.

It is important to note that in the equations and scales above, how each success and failure factor is defined can have an impact on its performance in the combined analysis. For example, if two similar factors were combined into a single factor, they may be considered more important together than individually. The identified success and failure factors are defined and discussed in the following sub-sections (5.2.2 and 5.2.3). However, a worked example of the above scales and equations for frequency, weighting and the combined analysis is first presented. The success factor, regulatory frameworks, is used as the example. As the calculation of overall success factors is averaged across policies and countries/states and is not related to a given policy or country/state, an example from section 5.2.5 is used (performance standards).

Importance of Regulatory Frameworks for Performance Standards:

- Success Factor: Regulatory frameworks (RF)
- Success Factor Frequency: 12 (above ≥5 frequency threshold)
- Success Factor Weighting: 2.8 (in 1.5-2.4 'some importance' weighting group)
- Policy Frequency: 81 (number of PS evaluations in the total sample of 690 evaluations)
- Policy Weighting: 4.2 (averaged policy weighting across evaluations in the sample)
- Combined Analysis: 8.2% (in 5.0-9.9% range so considered 'important'):

1) Frequency-Weighting combined analysis (FW_{com}) = Policy Success weighting (PS_w) x (Policy Success factor frequency in sample (PS_{ffs}) x Policy Success factor weighting (PS_{fw})) / 10

   14.1 = 4.2 x (12 x 2.8) / 10

2) Theoretical Maximum combined analysis Score (FWS_{commax}) = Policy Success weighting (PS_w) x (Policy frequency in sample (P_{fs}) x Theoretical Maximum Policy Success Weighting (PS_{wmax})) / 10

   171.4 = 4.2 x (81 x 5.0) /10

3) Frequency-Weighting combined analysis Score (FWS_{com}) = (Frequency-Weighting combined analysis (FW_{com}) / Theoretical Maximum combined analysis Score (FWS_{commax})) x 100%

   8.2% = (14.1 / 171.4) x 100%

In summary, regulatory frameworks is considered an 'important' factor for performance standards. This process was undertaken for all of the identified success and failure factors for each of the twelve main categories of DSM policy in sub-section 5.2.5 and for each of the 66 countries/states in sub-section 5.2.6.

5.2.2 Key Overall Success Factors

Success factors are those that contribute to the success of a DSM policy and represent the underlying mechanisms for policy success. The factors were determined inductively from the 119 documents; they were not pre-determined. The saturation point at which no new success factors became evident in the analysis was ~50 documents in the systematic review. The purpose of continuing the systematic review past ~50 documents was due to the nature of systematic reviews. Unlike qualitative interviews, systematic reviews do not aim to reach saturation points in the determination of sample sizes. Instead, the sample size is established by the pre-determined databases that need to be consulted. As discussed in chapter three, a quick scoping review should be conducted prior to a systematic review in order to identify the key databases that publish research on the topic of interest. Once the databases have been identified, the systematic review is not complete until all of the databases have been synthesised. Thus, the sampling strategy for systematic reviews is non-probability and purposive sampling, as the population of interest is the total number of documents (initial hits) from the identified databases and the sample is the number of relevant documents that pass stage three of the systematic review protocol (the inclusion/exclusion criteria stage).

Although the MCDM analysis is not analysed and discussed separately from the systematic review, it is important to note that the same principles apply to the MCDM interviews. The quantitative nature of MCDM analysis involves the sample size usually being determined by what is required for statistical analysis rather than by reaching a saturation point. However, in this research, the sample size was determined neither by statistical sample sizes nor saturation points. Instead, the method used the same sampling strategy as that of the systematic review (non-probability, purposive sampling). The literature review identified the main experts from academia, industry and policy and these were contacted by email. The overall response rate was 52.6% (20/38 responded) and the overall participation rate was 44.7% (17/38 participated). As such, the number of interviews was determined by the number of experts that could be interviewed (the sample) from the total number of experts identified (the population of interest). The approach is justified for two reasons: firstly, it allows the MCDM interviews and the systematic review to be more closely aligned, which is important in the triangulation of the two methods; and secondly, the MCDM interviews are a much smaller part of the research, which were inputted into the systematic review as one of the 35 databases.

Following the synthesis of the 35 databases, a list of 22 success factors were identified overall (across DSM policies and countries/states). These can be broken down into five broad categories: regulatory support, financial support, policy characteristics, stakeholder engagement and infrastructure. However, the analysis takes place at the success factor level rather than the category level. Although all 22 of the identified success factors are important for DSM policies to succeed, the purpose of the analysis was to determine which factors are more important than others in three levels of analysis: overall, by policy and by country/state.

Figure 30 presents the results of the frequency analysis, which are averaged across DSM policies and countries/states. The total number of documents (not evaluations) that discuss each factor is shown on the *y*-axis and the success factors are shown on the *x*-axis. The same pattern is produced if the total number of evaluations (rather than documents) is presented.

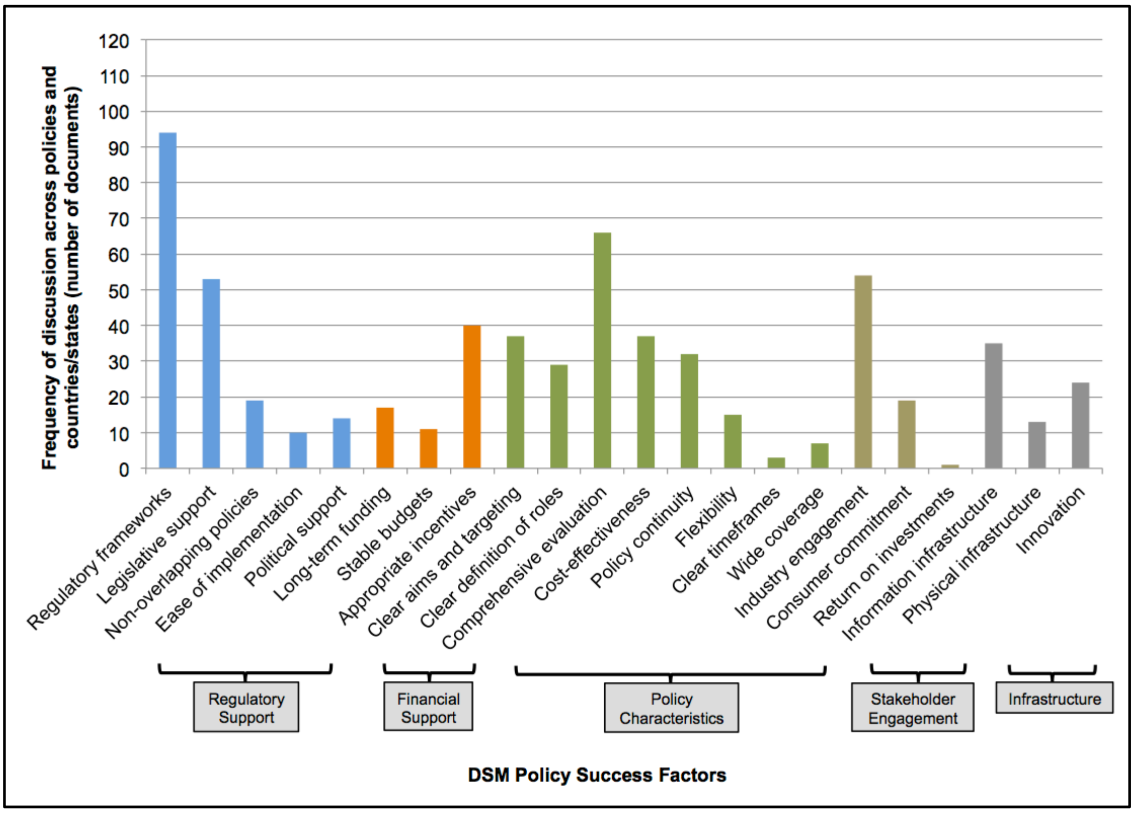


Figure 30: the overall frequency of DSM policy success factors

The research aims to collate and synthesise the results from the individual evaluations of specific policies in specific countries/states in order to reach generality and representation in patterns for success factors by policy and by country/state. Figure 30 shows that five factors are related to regulatory support, three factors are related to financial support, eight factors are related to policy characteristics, three factors are related to stakeholder engagement and three factors are related to infrastructure. Table 12 gives an explanation of what each success factor refers to.

Success Factor Category	Success Factor	Success Factor Explanation
Regulatory support	Regulatory frameworks (RF)	Regulatory rules, government orders, policy frameworks, and policy guidance
Regulatory support	Legislative support (LS)	Legislation, legal rules, and penalties
Regulatory support	Non-overlapping policies (NO)	Policies complement each other but do not overlap, policies are joined-up and integrated
Regulatory support	Ease of implementation (EI)	Politically feasible, practically feasible within the dedicated resources and timescales
Regulatory support	Political support (PS)	Cross-party support, support from senior ministers or commissioners
Financial	Long-term funding	Funding provided for long-term continuation of the

support	(LF)	policy, policy is re-funded in each policy period, budgets may fluctuate but are long-term
Financial support	Stable budgets (SB)	Budgets are stable and do not fluctuate over the policy period, budgets are not necessarily long-term
Financial support	Appropriate incentives (AI)	Well-designed incentives, which are appropriate to the targeted party, incentives may be financial or other
Policy characteristics	Clear aims and targeting (CA)	Policy aims, objectives, and targets are clearly defined and communicated to relevant parties, policy is appropriately targeted
Policy characteristics	Clear definition of roles (CR)	Policy clearly defines and communicates the roles of relevant parties and stakeholders
Policy characteristics	Comprehensive evaluation (EV)	Enough resources are dedicated to the comprehensive evaluation of policy design, implementation, and monitoring
Policy characteristics	Cost-effectiveness (CE)	Policy is cost-effective for all parties involved (not just the government), policy is cost-efficient
Policy characteristics	Policy continuity (PC)	Policy is continued over the long-term across successive governments where relevant, excessive changes are not made during policy periods (though changes may be made between policy periods)
Policy characteristics	Flexibility (FX)	Policy has some flexibility in making changes where policies are failing (poor designs or unpopular with relevant parties), some DSM policies benefit more from greater flexibility than others (e.g. voluntary policies compared with utility obligations)
Policy characteristics	Clear timeframes (CT)	Timeframes for policy are clearly defined and communicated to relevant parties, timescales for policy are appropriate
Policy characteristics	Wide coverage (WC)	Policy reaches a large number of the targeted population, policy can be wide ranging in terms of sectors covered or in terms of the total number of consumers/stakeholders reached within a targeted sector (or both)
Stakeholder engagement	Industry engagement (IE)	All relevant stakeholders that are not consumers are appropriately engaged, commitment from relevant parties in industry
Stakeholder engagement	Consumer commitment (CC)	All targeted consumers are appropriately engaged, commitment from the national or state population (not just those directly targeted)
Stakeholder engagement	Return on investments (RI)	All relevant stakeholders receive adequate returns on investments, primarily refers to utilities engaging in DSM programmes but can also include returns to governments and consumers where relevant
Infrastructure	Information infrastructure (IF)	Policies are clearly communicated through appropriate information channels, relevant parties are appropriately educated on the policy
Infrastructure	Physical infrastructure (PI)	Required physical infrastructure (where relevant) is provided for the policy to be implemented successfully
Infrastructure	Innovation (IV)	Policy stimulates innovation in DSM such as new technologies, new processes, and new techniques

Table 12: explanations of DSM policy success factors

It is clear from the frequency analysis that regulatory support, particularly in the form of regulatory frameworks, is the most frequently discussed factor and was present in almost all of the documents in the systematic review. For regulatory frameworks, 94/119 documents discuss the factor (79.0% of the sample). Comprehensive evaluation, industry engagement and legislative support are also frequently discussed in 66/119, 54/119 and 53/119 documents respectively (55.5%, 45.4% and 44.5% of the sample respectively). At the other end of the scale, return on investments (1/119), clear timeframes (3/119) and wide coverage (7/119) were only mentioned in a few evaluations (0.8%, 2.5% and 5.9% of the sample respectively).

Figure 31 uses the same structure, factor grouping, and colour-coding as figure 30 to show the results of the weighting analysis. The figure presents the averaged weighting of each factor on the *y*-axis, and the success factors on the *x*-axis. Although the weighting scale runs from 1-3 (to one decimal place), it is possible for factors to score <1.0 if some of the individual evaluations do not assign weightings to various factors (thus considering them unimportant). This is taken into account in the calculation of the average weightings.

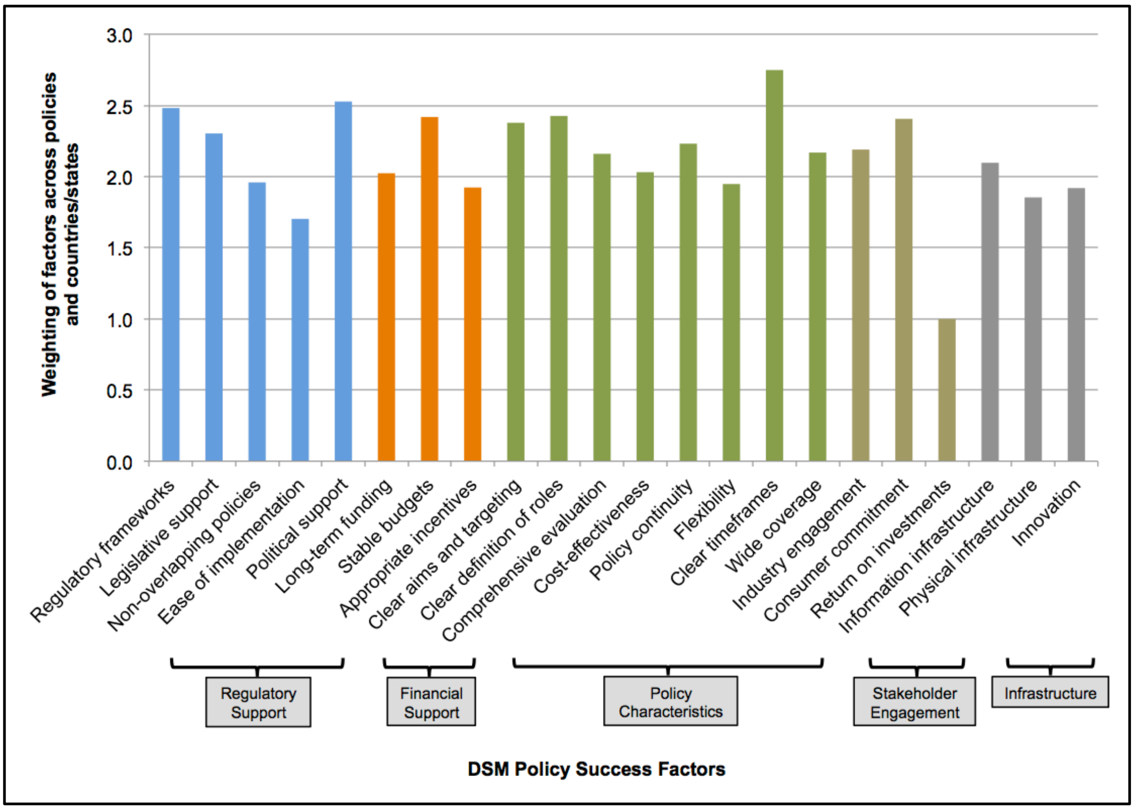


Figure 31: the overall weighting of DSM policy success factors

Figure 31 highlights that clear timeframes has the highest weighting (2.8), followed by regulatory frameworks (2.5) and political support (2.5). At the other end of the scale, return on investments (1.0 weighting) and ease of implementation (1.7 weighting) have the lowest weightings. The results convey the dominance of regulatory frameworks as a crucial factor in both the frequency and weighting analyses, and the low scoring of return on investments and ease of implementation in both analyses. However, the results show that clear timeframes is considered one of the most important factors in the weighting analysis, despite being one of the least discussed factors in the frequency analysis. The same applies to political support (highly weighted but not frequent). This justifies the use of the combined analysis in sub-sections 5.2.5 and 5.2.6.

5.2.3 Key Overall Failure Factors

Failure factors represent the underlying mechanisms for DSM policy failure. Using the same analytical process as for policy success, the factors were determined inductively from the 119 documents and were not pre-determined. 25 failure factors were identified, which can be broken down into five broad categories: policy issues, consumer issues, political issues, industry issues and infrastructure issues. Figure 32 summarises the failure factors and their frequency of discussion in the sample. The figure uses the same structure as figure 30.

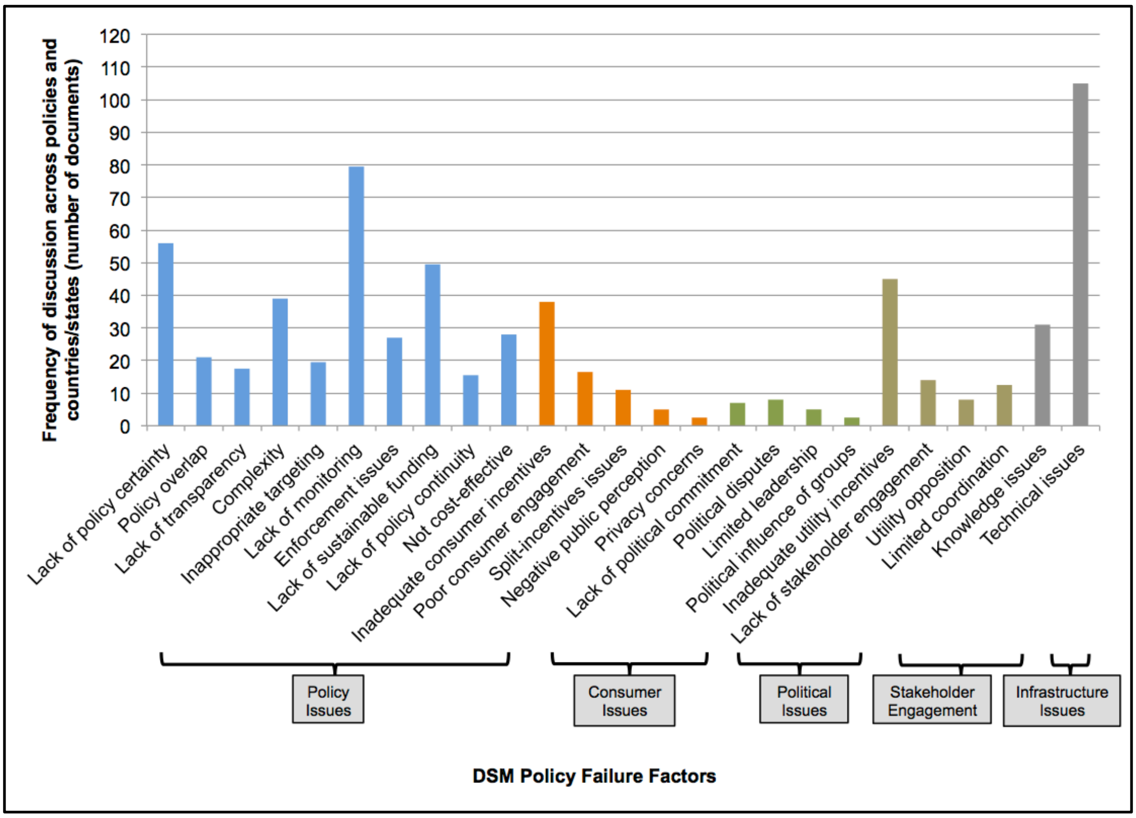


Figure 32: the overall frequency of DSM policy failure factors

Figure 32 shows that ten factors are related to policy issues, five factors are related to consumer issues, four factors are related to political issues, four factors are related to stakeholder engagement and two factors are related to infrastructure issues. Table 13 gives an explanation of what each failure factor refers to.

What is clear from the frequency analysis is that technical issues, which refers primarily to programme management and administration issues, but also to technological performance problems, is the most frequently discussed factor and was present in almost all of the documents in the systematic review (105/119 documents – 88.2% of the sample). A lack of monitoring (80/119 documents – 67.2% of the sample), a lack of policy certainty (56/119 documents – 47.1% of the sample) and a lack of sustainable funding (50/119 documents – 42.0% of the sample) also rank highly. At the other end of the scale, privacy concerns and the political influence of interest groups were only mentioned in a few documents (3/119 documents each – 2.5% of the sample).

The weighting of failure factors was also examined and the results are presented in figure 33, which follows the same structure as figure 31.

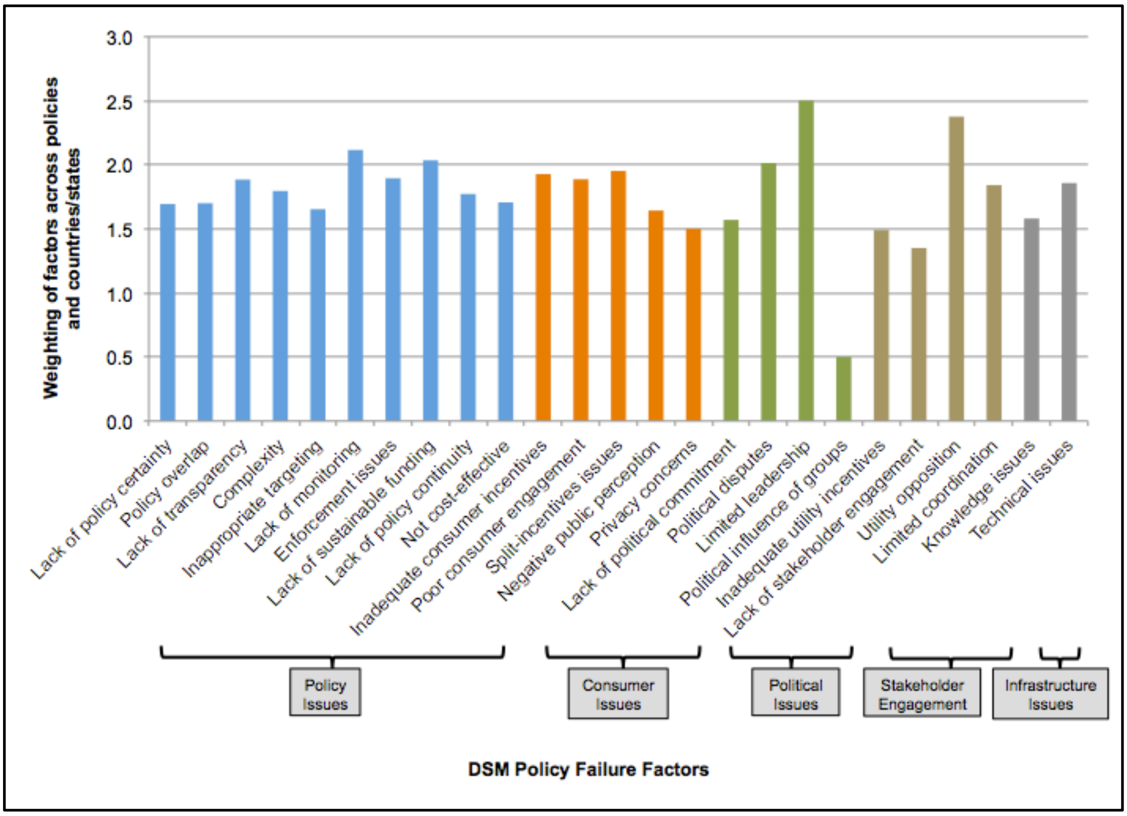


Figure 33: the overall weighting of DSM policy failure factors

Failure Factor Category	Failure Factor	Failure Factor Explanation
Policy issues	Lack of policy certainty (LC)	Unclear signals from government over policy details, stakeholder confusion due to policy uncertainty
Policy issues	Policy overlap (PO)	Policies do not complement each other and overlap, policies are not joined-up and integrated
Policy issues	Lack of transparency (LT)	Policy details are not transparent and are not clearly communicated to relevant parties
Policy issues	Complexity (CX)	Complex policies that place large administrative/financial burdens on relevant parties, stakeholder confusion due to complex policy requirements
Policy issues	Inappropriate targeting (IT)	Policy is not targeted to the most appropriate sector or does not adequately capture and engage the targeted sector (or both)
Policy issues	Lack of monitoring (LM)	Limited resources are dedicated to policy evaluation, limited monitoring conducted during the implementation and post-policy stages
Policy issues	Enforcement issues (EF)	Lack of enforcement of relevant policies, no formal penalties are present or penalties are not large enough to impact compliance levels
Policy issues	Lack of sustainable funding (SF)	Funding is not consistent and long-term, budgets fluctuate within and/or between parliamentary terms
Policy issues	Lack of policy continuity (CY)	Policy is not continued over the long-term across successive governments where relevant, excessive changes are made during and/or between policy periods
Policy issues	Not cost-effective (CF)	Policy is not cost-effective for all parties involved (not just the government), policy is not cost-efficient
Consumer issues	Inadequate consumer incentives (IC)	Poorly designed incentives which are not appropriate for targeted consumers and fail to incentivise them to undertake DSM activities
Consumer issues	Poor consumer engagement (PE)	Consumers are not adequately engaged, lack of communication and education, poorly designed and targeted communication
Consumer issues	Split-incentives issues (SI)	The differing incentives for landlords and tenants are not adequately addressed, inappropriate (or lack of) incentives for landlords and tenants
Consumer issues	Negative public perception (NP)	Policy is negatively perceived by the general population and/or targeted sector, negative communication of policy in the media
Consumer issues	Privacy concerns (PV)	Lack of data privacy for relevant consumers for some policies (e.g. smart meter rollouts), privacy issues during installation of energy efficiency measures
Political issues	Lack of political commitment (LP)	Lack of cross-party commitment, lack of support from senior ministers or commissioners, political support is not maintained over the long-term
Political issues	Political disputes (PD)	Disputes between government and relevant stakeholders, disputes between different political parties
Political issues	Limited leadership (LL)	Government is not willing to pioneer the implementation of some policies before other governments have done so, limited innovation in developing new policies
Political issues	Political influence of groups (IG)	Strong influence of interest groups against policy implementation, reduction in policy ambition due to strong lobbying influence of targeted parties
Industry issues	Inadequate utility	Poorly designed incentives which are not

	incentives (IU)	appropriate for targeted utilities and fail to incentivise them to undertake DSM activities
Industry issues	Lack of stakeholder engagement (LE)	Stakeholders are not adequately engaged in all stages of the policy process, lack of communication and consultation
Industry issues	Utility opposition (UO)	Utility disputes over level of policy ambition and administrative/financial resources required, strong lobbying influence of utilities against policy implementation
Industry issues	Limited coordination (CP)	Lack of coordination between government and relevant parties, lack of coordination between relevant parties
Infrastructure issues	Knowledge issues (KI)	General lack of awareness and understanding of policy, relevant parties not well-informed and educated on policy, poor communication of policy using inappropriate channels
Infrastructure issues	Technical issues (TI)	Programme administration issues for relevant parties and government, technological performance problems, lack of required physical infrastructure (where relevant)

Table 13: explanations of DSM policy failure factors

It is crucial to note that all 25 of the factors are important in determining the failure of a policy and other factors that are not included in the list are not considered to be important (i.e. factors not discussed in the systematic review). However, under the resource constraints of governments, the research identifies the factors where policy-makers should concentrate their limited resources.

5.2.4 Statistical Associations between Factors

An interesting extension to the main analysis is to explore whether or not there are significant associations between success factors and between failure factors. *Pearson's product-moment correlation coefficient* (*r*) is used, as the test examines associations between variables using the equation shown overleaf.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Equation 3: Pearson's correlation coefficient (*r*)

<u>*Key:*</u>

r = Pearson's correlation coefficient

n = number of pairs of scores

Σxy = sum of the products of paired scores

Σx = sum of x scores

Σy = sum of y scores

Σx^2 = sum of squared x scores

Σy^2 = sum of squared y scores

The purpose of the analysis is to test the significance of the associations between different success or failure factors in order to identify the clusters of factors that need to be present for DSM policies to be successful overall. The analysis uses the data from the frequency analysis and follows the premise that, as the frequency of one success (or failure) factor increases, another success (or failure) factor similarly increases. The degree of association is the primary focus and where two factors are significant at the 0.05 significance level, it is assumed that the success or failure of a policy will be determined by attention being given to both factors in the design and implementation of the policy. The 0.05 significance level is a standardised threshold in statistics, which states that there is <5% chance that the observed results have arisen purely by chance (Cochrane Handbook, 2012).

For success factors, each of the 22 factors in table twelve were correlated against each other based on the frequencies of those factors for each DSM policy in the 119 documents. For example, the frequency of discussion of regulatory frameworks for each of the twelve DSM policies was correlated against the frequency of discussion of appropriate incentives for each of the twelve policies, producing an *r* value of 0.92, which represents a strong positive association between the two success factors.

With 22 success factors, 20 degrees of freedom were necessary (*n* – 2) and an *r* value of 0.423 or greater was needed to test for significance. The results are presented in the 22 graphs (one for each success factor) included in the Appendix (Appendix Figure 4). The success factor acronyms are used to make the graphs more presentable, which are explained in table twelve. The dividing line is set at 0.423 (the significance level) where the *r* values of the success factors to the right of it (light grey) are significantly associated with the success factor in question, and those to the left of it (dark grey) are not. As a positive relationship or no relationship is expected, all negative correlations are disregarded, as theoretically, the absence of certain success factors should not increase the probability that a policy will succeed, but increase the probability that the policy will fail or have no impact. No graph shows a significant negative relationship (past the -0.423 threshold), which confirms the expectation of a positive relationship. The one exception to this is the weak negative association between clear timeframes and consumer commitment, which produces an *r* value of -0.44. By examining the raw data, it is clear that this is due to the low number of data points that exist for clear timeframes (three data points), which produces a situation where differences between the three data points and their equivalents for consumer commitment results in a disproportionately larger change in the *r* value compared with the rest of the factor associations where the sample sizes are much larger. Thus, the finding should be rejected due to the small sample size.

The graphs show that legislative support (LS) and regulatory frameworks (RF) have the greatest number of significant associations with other success factors (14 and 13 other factors respectively). This highlights that in order to increase the probability that a DSM policy is successful, legislative support and regulatory frameworks need to be present. The top success factors by the number of associations with (>5) other factors are listed overleaf.

- Legislative support (LS) (14 factor associations)
- Regulatory frameworks (RF) (13 factor associations)
- Non-overlapping policies (NO) (11 factor associations)
- Appropriate incentives (AI) (10 factor associations)
- Clear aims and targeting (CA) (10 factor associations)
- Clear definition of roles (CR) (8 factor associations)
- Ease of implementation (EI) (7 factor associations)
- Comprehensive evaluation (EV) (7 factor associations)
- Cost-effectiveness (CE) (6 factor associations)

The importance of a factor in this respect can be presented in terms of the proportion (%) of other success factors that it is significantly associated with. The thesis proposes a scale to visualise this, as shown in figure 34:

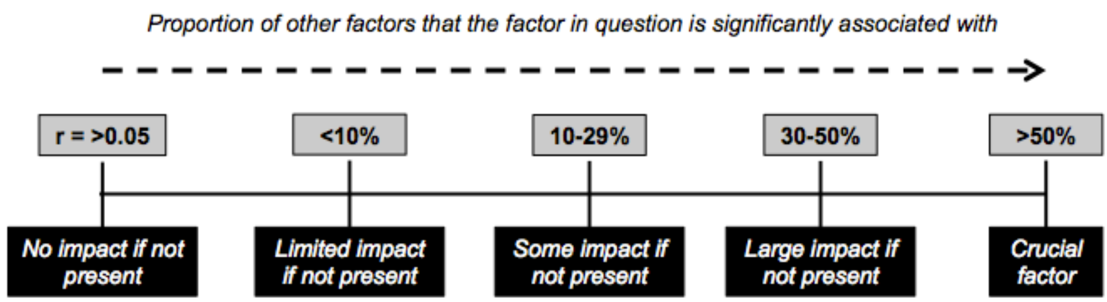


Figure 34: the *Success Factor Association Scale* to show the percentage of the sample that is significantly associated with the factor in question

The *Success Factor Association Scale* is designed in a similar way to the five-part *Policy Success Weighting Scale* (shown in figure 28) in order to provide standardisation between the scales used in the analysis for research question two. The *Success Factor Association Scale* was applied to all 22 of the success factors. Legislative support and regulatory frameworks are ‘crucial’ factors, being associated with 63.6% and 59.1% of the sample respectively. Thus, they fit into the category on the far right of the scale. Worked examples of how the percentages were derived are shown below. As with the discussions in the previous sub-sections, how each factor is defined can have an impact on the results.

Equation:

(Number of factor associations / Total number of other factors) x 100%

Legislative Support (LS): (14 / 22) x 100% = 63.6% of the sample

Regulatory Frameworks (RF): (13 / 22) x 100% = 59.1% of the sample

The percentages for the success factors with the greatest numbers of associations with other factors are given below:

- Legislative support (LS) (63.6% of sample)
- Regulatory frameworks (RF) (59.1% of sample)
- Non-overlapping policies (NO) (50.0% of sample)
- Appropriate incentives (AI) (45.5% of sample)
- Clear aims and targeting (CA) (45.5% of sample)
- Clear definition of roles (CR) (36.4% of sample)
- Ease of implementation (EI) (31.8% of sample)
- Comprehensive evaluation (EV) (31.8% of sample)
- Cost-effectiveness (CE) (27.3% of sample)

NO, AI, CA, EI and EV appear to have a large impact if not present and thus fit into the category that is second from the far right on the scale. CE fits into the middle category on the scale and has some impact if not present. Other success factors that fit into this category are shown below:

- Political support (PS) (3 factor associations – 13.6% of sample)
- Policy continuity (PC) (3 factor associations – 13.6% of sample)
- Wide coverage (WC) (3 factor associations – 13.6% of sample)
- Industry engagement (IE) (3 factor associations – 13.6% of sample)

The following factors have a limited impact if not present and fit into the category that is second from the far left on the scale:

- Clear timeframes (CT) (2 factor associations – 9.1% of sample)
- Consumer commitment (CC) (2 factor associations – 9.1% of sample)
- Information infrastructure (IF) (2 factor associations – 9.1% of sample)
- Flexibility (FX) (1 factor association – 4.6% of sample)

- Physical infrastructure (PI) (1 factor association – 4.6% of sample)

Finally, the following factors show no significant associations with any of the 21 other success factors. As such, they fit into the category on the far left of the scale (0.05 is the standardised significance threshold, as discussed above):

- Long-term funding (LT) (0 factor associations – 0.0% of sample)
- Stable budgets (SB) (0 factor associations – 0.0% of sample)
- Return on investments (RI) (0 factor associations – 0.0% of sample)
- Innovation (IV) (0 factor associations – 0.0% of sample)

For failure factors, the same process was undertaken. The 25 factors in table 13 were correlated against each other based on the frequencies of those factors for each DSM policy in the 119 documents. With 25 failure factors, 23 degrees of freedom were necessary (n-2) and an *r* value of 0.396 or greater was needed to test for significance. The results are presented in the 25 graphs (one for each failure factor) shown in the Appendix (Appendix Figure 5). As with the success factor graphs, the failure factor acronyms are similarly used to make the graphs more presentable, which are explained in table thirteen. The dividing line is set at 0.396 (the significance level) where the *r* values of the failure factors to the right of it (light grey) are associated with the failure factor in question, and those to the left of it (dark grey) are not (as per Appendix Figure 4 for success factors).

The graphs show that the failure factors below have the greatest number of associations with (>5) other failure factors (the percentages are included using the same calculations as per figure 34):

- Policy overlap (PO) (15 associations – 60.0% of sample)
- Lack of transparency (LT) (14 associations – 56.0% of sample)
- Complexity (CX) (13 associations – 52.0% of sample)
- Lack of monitoring (LM) (13 associations – 52.0% of sample)
- Inappropriate targeting (IT) (8 associations – 32.0% of sample)
- Lack of sustainable funding (SF) (8 associations – 32.0% of sample)
- Lack of policy continuity (CY) (8 associations – 32.0% of sample)

- Enforcement issues (EF) (7 associations – 28.0% of sample)
- Not cost-effective (CF) (7 associations – 28.0% of sample)
- Inadequate consumer incentives (IC) (6 associations – 24.0% of sample)
- Poor consumer engagement (PE) (6 associations – 24.0% of sample)
- Split-incentives issues (SI) (6 associations – 24.0% of sample)

If the failure factors are placed in the *Success Factor Association Scale*, PO, LT, CX and LM are 'crucial' factors, and IT, SF and CY have a large impact if not present. In addition to EF, CF, IC, PE and SI, the following failure factors have some impact if not present:

- Inadequate utility incentives (IU) (4 associations – 16.0% of sample)
- Privacy concerns (PV) (3 associations – 12.0% of sample)
- Political disputes (PD) (3 associations – 12.0% of sample)
- Political influence of groups (IG) (3 associations – 12.0% of sample)

The following failure factors have a limited impact if not present:

- Negative public perception (NP) (2 associations – 8.0% of sample)
- Limited leadership (LL) (2 associations – 8.0% of sample)
- Lack of stakeholder engagement (LE) (2 associations – 8.0% of sample)
- Limited coordination (CP) (2 associations – 8.0% of sample)
- Lack of political commitment (LP) (1 association – 4.0% of sample)
- Knowledge issues (KI) (1 association – 4.0% of sample)

The final three factors, which are listed below, show no significant associations with any of the 24 other failure factors:

- Lack of policy certainty (LC) (0 associations – 0.0% of sample)
- Utility opposition (UO) (0 associations – 0.0% of sample)
- Technical issues (TI) (0 associations – 0.0% of sample)

It is important to reiterate that the *Success Factor Association Scale* does not rank factors individually by importance – this is the purpose of the combined frequency-weighting analysis. Instead, the scale identifies the factors that are important only in the context of specific other factors being present. For example, technical issues (primarily programme administration problems) is the most frequently discussed failure factor in the systematic review sample and has a reasonably high weighting. The finding that its presence in causing policy failure is not determined by the presence of other factors highlights the crucial need to overcome this factor in the design and implementation of DSM policies. However, the image is different for success factors. The importance of regulatory frameworks as the most frequently discussed factor in the systematic review sample and one of the most highly weighted factors, is coupled with the need for it to be present alongside thirteen other success factors to ensure a DSM policy is successful. Despite this, the results are more fruitful for answering research question two when the success and failure factors are broken down by DSM policy (sub-section 5.2.5) and by country/state (sub-section 5.2.6).

5.2.5 Key Success and Failure Factors by Policy

This sub-section details the key success and failure factors by DSM policy. It concentrates on the twelve individual policies discussed in the previous chapters:

- Incentive payment-based demand response (IPBDR)
- Price-based demand response (PBDR)
- Market transformations (MT)
- Infrastructure rollouts (IR)
- Utility obligations (UO)
- Labelling (LB)
- Performance standards (PS)
- Loans and subsidies (L&S)
- Utility business models (UBM)
- Research and development programmes (R&D)
- Information campaigns (IC)
- Voluntary programmes (VP)

Due to the limited number of evaluations that examine policy packages, these are excluded from this part of the analysis. The same analytical process to the analysis of the overall success and failure factors (across DSM policies and countries/states) is used in this section. The main results from the combined frequency-weighting analysis are discussed, though the results from the individual frequency and weighting analyses are also graphically summarised in figures 35 and 36. The twelve DSM policies are examined in turn, and the calculations for frequency, weighting, and the combined analysis are given for the first discussed policy (incentive payment-based demand response) to show how the results were derived. The same calculation process was used for all twelve of the DSM policies in the research.

Incentive Payment-Based Demand Response

Incentive payment-based demand response (IPBDR) refers to tariffs that encourage the reduction or shifting of load, particularly during peak times. From the combined analysis, regulatory frameworks was the only success factor to pass the threshold, and technical issues and a lack of policy certainty were the key failure factors.

Success Factor: Regulatory frameworks (RF)
Success Factor Frequency: 13 (above ≥5 frequency threshold)
Success Factor Weighting: 2.3 (in 1.5-2.4 ‘some importance’ weighting group)
Policy Frequency: 62 (number of IPBDR evaluations in the sample)
Policy Weighting: 3.4 (averaged policy weighting across evaluations in sample)
Factor Combined Analysis Score: 3.4 x (13 x 2.3) / 10 = 10.5
Policy Theoretical Maximum Score: 3.4 x (62 x 5.0) / 10 = 106.4
Combined Analysis Percentage Score: (10.5 / 106.4) x 100% = 9.8%

Thus, regulatory frameworks is considered the most important success factor for IPBDR, as the combined analysis score as a percentage is above the 5.0% threshold and falls into the 5.0-9.9% ‘important’ group (as stated before, ‘crucial’ factors score ≥10.0% and ‘unimportant’ factors score <5.0%).

Failure Factor: Technical issues (TI)

Failure Factor Frequency: 8 (above ≥5 frequency threshold)

Failure Factor Weighting: 2.3 (in 1.5-2.4 ‘some importance’ weighting group)

Policy Frequency: 62

Policy Weighting: 3.4

Factor Combined Analysis Score: 3.4 x (8 x 2.3) / 10 = 6.4

Policy Theoretical Maximum Score: 3.4 x (62 x 5) / 10 = 106.4

Combined Analysis Percentage Score: (6.4 / 106.4) x 100% = 6.0%

The calculations show that technical issues is considered one of the two most important failure factors for IPBDR, as the combined analysis score as a percentage is above the 5.0% threshold (and falls into the 5.0-9.9% ‘important’ group). A similar score is produced for a lack of policy certainty (6.3%):

Failure Factor: Lack of policy certainty (LC)

Failure Factor Frequency: 8

Failure Factor Weighting: 2.4

Policy Frequency: 62

Policy Weighting: 3.4

Factor Combined Analysis Score: 3.4 x (8 x 2.4) / 10 = 6.7

Policy Theoretical Maximum Score: 3.4 x (62 x 5) / 10 = 106.4

Combined Analysis Percentage Score: (6.7 / 106.4) x 100% = 6.3%

For success factors, regulatory frameworks allow the introduction of IPBDR in balancing or capacity markets, as they establish the rules and processes for participation. In the UK, the inclusion of permanent electricity reductions through energy efficiency as a new resource in the (newly created) capacity market (which had its first auction in December 2014) was piloted for two years from June 2014 (UK DECC, 2014c). A key aim of the pilot is to develop the required rules and processes to encourage a greater contribution from IPBDR. The PJM Interchange in the USA represents one of the few capacity markets globally to include energy efficiency as a resource and this was implemented in 2009 (Titus *et al.*, 2009). The UK’s pilot aims to learn from this experience. Although IPBDR is implemented at a regional level in the USA through system operators, Cappers *et al*. (2010) highlight the political interest at a national level to assess its size and scope in the country. The *Energy Policy Act of 2005* directed the Federal Energy Regulatory Commission (FERC) to do this (Cappers *et al*., 2010). Thus, it is clear from the evaluations that regulatory frameworks are essential for IPBDR to be successfully implemented and developed.

For failure factors, the findings highlight the challenges of successfully introducing the necessary regulations for IPBDR. Technical issues, such as programme administration problems and the risk of consumers not responding during peak times (as the penalty for non-response can be high for utilities during peak times), have a defining impact on determining failure. For example, in the summer of 2007 in the ISO New England ancillary services market in the USA, there was ~50% commitment from all demand response resources and despite ~45 minutes warning prior to the end of peak events, both demand response and generation resources took 20 minutes to return to baseline levels (Agnew *et al*., 2007). In the US state of Texas, the restructuring of electricity markets in 2002 led to a range of technical difficulties in continuing demand response programmes due to the need to deal with many retail electric providers (Zarnikau, 2010). Restructuring led to extensive measurement and verification requirements and problems in securing compensation from the retail electric providers (Zarnikau, 2010).

Furthermore, a lack of policy certainty and clarity makes it difficult for traditional or new utilities to invest in IPBDR. For example, in the UK, there was a lack of clarity over the continuity of incentives for the participation of demand response in the capacity market, which reduced participation (UKDRA MCDM interview, 2014). For demand response, contract lengths are currently limited to just one year (UKDRA MCDM interview, 2014).

Price-Based Demand Response

Price-based demand response (PBDR) refers to tariffs that vary the price of electricity or gas based on the time-of-day or the time-of-year. Such policies aim to increase the costs of consuming energy during peak times above the ‘normal’ rate and reduce the costs of energy during off-peak periods below the ‘normal’ rate (such as during the night). Thus, the focus of PBDR is on reducing or shifting load due to energy price changes, whereas the focus of IPBDR is on reducing or shifting load due to direct financial payments (Albadi and El-Saadany, 2008). From the combined analysis, regulatory frameworks, appropriate incentives, information infrastructure, clear aims and targeting and consumer commitment were the key success factors to pass the threshold, and technical issues, a lack of monitoring, inadequate utility incentives and inadequate consumer incentives were the key failure factors.

For success factors, the findings show the similarity between PBDR and IPBDR with regards to the importance of regulatory frameworks in the implementation of demand response resources. The main difference is the greater number of success factors for PBDR, which is partly explained by the increased complexity of PBDR policies and regulations compared with those for IPBDR. Time-of-use tariffs are usually more challenging to implement and administer than direct contracts between large consumers and the system operator. Time-of-use tariffs are usually implemented by energy suppliers for whole customer segments, such as the residential or commercial sectors (though some large consumers also have time-of-use tariffs). Numerous evaluations in the sample showed that regulatory frameworks are essential for PBDR to be successful, and the finding was replicated across countries/states and contexts (for example, Agnew *et al*., 2009, which looked at the New England region in the USA; Cappers *et al*., 2010, which looked at the USA; Wang *et al*., 2010, which looked at China).

During short-term energy crises, the evidence base shows that both categories of demand response (PBDR and IPBDR) can be particularly useful. For example, Lowry *et al*. (2004) evaluated the contribution of demand response in the 2000-2001 electricity crisis in the US state of California and found that the appropriateness of the incentives needed to be matched with educating consumers on the issues and how they can benefit from contributing to overcoming them. For instance, it was important that customers were paid for the capability to reduce consumption instead of just taking into account actual reductions (Lowry *et al*., 2004). This applies to both categories of demand response, though appears more prominently in the sample as a factor for PBDR.

Linked to this, Walawalkar *et al*. (2010) show the importance of consumer commitment by highlighting that, despite the need for only modest requirements from consumers, success depends on the consumer acceptance of altering the use of equipment and systems (lifestyle and comfort changes) or changes in operating procedures. Thus, it is crucial that the policy is well-targeted and has clear requirements. In China, large-scale PBDR pilots were undertaken at a provincial level in 2003 and were targeted to specific sectors. In Beijing and Jiangsu, residential time-of-use pricing was implemented (the on-peak price in Jiangsu was US $66.51/MWh and the off-peak price was US $36.28/MWh, compared with US $66.88/MWh for those not on the tariff) and in Guangdong, industrial time-of-use pricing was implemented (which reduced load by 500 MW in 2003) (Wang *et al*., 2010). As such, these programmes were generally considered to be successful (Wang *et al*., 2010).

For failure factors, the findings show some similarities between PBDR and IPBDR, particularly in relation to technical issues as the most crucial factor. Most of the demand response evaluations focussed on the USA or China at a sub-national (state or provincial) level. For example, Wang *et al*. (2010) highlight how technical issues, such as reaching a wider range of sectors than just the industrial sector (the residential sector is only included for a few consumers in Beijing and Jiangsu), are important to overcome in China. Within the USA, Wirtshafter *et al*. (2007) examined small-scale consumers with time-of-use tariffs and found that the 20/20 Program in California (which ran over the summer of 2005) did not adequately incentivise utilities and utility costs were US $67 million. Furthermore, Wirtshafter *et al*. (2007) identified that 75% of the rebate dollars paid were to households that were not aware of the programme or were not actively trying to save energy in response to it. Inadequate consumer incentives and inadequate utility incentives appear to be equally important. For example, Walawalkar *et al*. (2010) found that in the US PJM and NYISO markets, economic-based demand response programmes have not provided sufficient revenues for customers to respond during off-peak hours.

Underlying the other failure factors for PBDR is a lack of monitoring and standardised reporting practices and metrics, which hinders reliability assessments (Cappers *et al*., 2010). Alongside technical issues, this was found to be one of the most crucial factors overall, across DSM policies and countries/states. However, in summary, ensuring that the right regulatory frameworks and incentives are in place are the most important factors determining the success or failure of PBDR.

Market Transformations

Market transformations (MT) refer to broad and long-term policies to overcome the market barriers to (primarily) energy efficiency development and to help stimulate market development. From the combined analysis, appropriate incentives was the only success factor to pass the threshold, and technical issues, a lack of monitoring and a lack of policy certainty were the key failure factors.

For success factors, the findings convey that for MT to be successful, long-term support is necessary, but this support needs to involve appropriate incentives. The evaluations of MT show that simply providing financial support is inadequate if it does not stimulate the development of DSM markets. For example, Wikler (2000) highlights Thailand as an example of a country that has successfully and appropriately established financial support for MT programmes – the country obtained funding from the World Bank and the Australian government to develop its MT programmes (Wikler, 2000). Colby and Davis (2011) identified a similar finding for the US state of New York. In 2010, ARRA (*American Recovery and Reinvestment Act of 2009*) funding provided individual rebates for the purchase of ENERGY STAR refrigerators (US $75), freezers (US $50) or clothes washers (US $75), or for an ENERGY STAR three-appliance bundle consisting of a CEE Tier 2 or 3 refrigerator and clothes washer and CEE Tier 1 or 2 level dishwasher (US $500). The *New York Energy Smart Products programme* (NYESP) promoted the rebates and the programme was put on hold twice when the New York State Energy Research Development Authority (NYSERDA) received enough applications to potentially use up all of the available funds (Colby and Davis, 2011). NYESP achieved a gradual increase in ENERGY STAR home appliance market shares for refrigerators, freezers and clothes washers (Colby and Davis, 2011), and the programme has generally been considered a success.

For failure factors, the findings show that the technical complexity and difficulty in monitoring the long-term effects of MT are the primary factors that contribute to policy failure. Furthermore, the long-term consistency and stability of policies appears crucial. Many of the successful market transformation experiences come from the USA. For example, Gillich and Sunikka-Blank (2013) examined the *Better Buildings Neighborhood Program*, Colby and Davis (2011) looked at relevant programmes within ARRA and McRae *et al*. (2011) analysed the 2005-2009 *BetterBricks* programme. However, McRae *et al*. (2011) highlight the difficulty in evaluating market transformations and present the example of the Northwest Energy Efficiency Alliance (NEEA) to show this. The paper portrays how the NEEA did not manage to establish a proper baseline for the final evaluation and consequently, pre-policy and post-policy differences between participants and non-participants could not be established. The paper gives examples of how evaluation could be done to overcome this issue, which revolves around a methodology that encourages communication and the development of consistent terminology with specific definitions that all parties can commonly understand. However, the evaluations highlight the challenges in evaluating long-term DSM policies such as MT.

Infrastructure Rollouts

Infrastructure rollouts (IR) refer primarily to smart meter rollouts (with or without energy display monitors). From the combined analysis, regulatory frameworks, legislative support, appropriate incentives, information infrastructure, industry engagement, consumer commitment, clear aims and targeting, political support, physical infrastructure, non-overlapping policies and policy continuity were the key success factors to pass the threshold, and technical issues, limited coordination, a lack of transparency, political disputes, a lack of policy certainty, inadequate utility incentives, limited leadership, policy overlap, privacy concerns, negative public perception, knowledge issues and a lack of policy continuity were the key failure factors.

For success factors, the full range of success factor types appears to be important (regulatory support, financial support, policy characteristics, stakeholder engagement and infrastructure, as per the categories in figure 30). IR requires clear regulation and legislation to establish the rules, aims, budgets and appropriate communication links that are required for the rollout (Kushler and Vine, 2003; Lowry *et al.*, 2004). The policy is designed to improve the physical infrastructure of meters in the premises of (primarily) small consumers and the technical aspects of the smart meter are important to establish. The UK's smart meter rollout between 2015-2020 includes an in-home display (IHD), which is an energy display monitor to visually show consumers how much energy (electricity and gas) they are using and other information, such as connection link quality, local time, active tariff price, cumulative consumption, customer identification number, debt, debt recovery rate, emergency credit balance, historic consumption, instantaneous active power import, low credit alert, meter balance, payment mode and power threshold status (see UK DECC, 2014b for the full technical specifications).

Additionally, it is clear that political support, industry engagement and consumer commitment are crucial, as the smart meter rollouts in the European Union (EU) have shown (University of Reading MCDM interview, 2014). Consumer engagement can be more effectively achieved if IR policies have clear aims, are well-targeted and provide appropriate incentives for consumers to reduce energy consumption. This is particularly evident in the EU, where IR policies have been most active. However, as these policies are currently being implemented (as a result of the *Smart Meter Rollout Directive* (Directive 2009/72/EC)), evaluations are unlikely to be available until after 2020 (the deadline for the completion of the rollouts). As such, the findings are based on a limited number of evaluations, as discussed in chapter four, and should be updated once more evidence becomes available.

For failure factors, the full range of failure factor types appears to be important (policy issues, consumer issues, political issues, stakeholder engagement and infrastructure issues, as per the categories in figure 32). Policy design issues and political disputes caused a number of EU countries to suffer delays in the rollout of smart meters, such as in the Netherlands (Cuijpers and Koops, 2012). Furthermore, where consumers are not properly incentivised (financial or non-financial) and do not accept the policy, implementation issues can occur. These issues occurred in the Netherlands where technical aspects were prioritised over the interests of the end users (Hoenkamp *et al.*, 2011), such as privacy issues (Cuijpers and Koops, 2012). This caused complications and legislative delays. Similar concerns were discussed in the design of the UK's smart meter rollout (McKenna *et al*., 2012), which pushed back the rollout period from 2014-2019 to 2015-2020. However, due to delays in the setting up of the Data Communications Company to deal with privacy and data concerns, it is likely that the rollout will not begin until early 2016 (UK DECC, 2014d). In addition to consumer concerns, it is clear from the evaluations that the utilities responsible for undertaking the rollout need to be adequately incentivised through cost-recovery mechanisms (or equivalent) and should be given clear instructions as to what is required of them.

Utility Obligations

Utility obligations (UO) refer to usually mandatory (though sometimes voluntary) obligations placed on suppliers, distributors, public entities, or building owners or users (THINK, 2012). Nevertheless, the evidence base is dominated by evaluations of obligations on energy suppliers and distributors. The obligations aim to meet various policy objectives, and there are a number of ways in which the targets can be expressed. The evidence base shows that targets are commonly set in terms of energy or carbon savings with sub-targets for fuel poverty (consumers living on a low income in a home that cannot be kept warm at a reasonable cost, as defined in the UK's *Warm Homes and Energy Conservation Act 2000* and reviewed in the Hills Fuel Poverty Review, 2012). From the combined analysis, regulatory frameworks, legislative support, comprehensive evaluation, clear definition of roles and cost-effectiveness were the key success factors to pass the threshold, and no failure factors passed the threshold. However, failure factors were produced in the individual frequency and weighting analyses, as summarised in figure 36. Nevertheless, only factors that are both frequent and highly weighted are discussed in this sub-section.

For success factors, although UO is arguably one of the more complex types of DSM policy to implement, it is becoming increasingly popular around the world as countries/states follow the successful experiences in the USA at a state-level and the UK. Both countries have had a long history of successfully implementing utility obligations (referred to as energy efficiency resource standards (EERS) in the USA and supplier obligations in the UK). The identified success factors highlight that UO requires more attention being given to the regulatory and policy support factors than to the other success factor categories (financial support, stakeholder engagement and infrastructure). In the sample, the UOs in the UK, the USA, Italy, France and Denmark were the most evaluated countries and different evaluations of the same policies were in agreement as to the success of the policies.

In the UK, the evaluations of supplier obligations since 2002 (*Energy Efficiency Commitment*, *Carbon Emissions Reduction Target* and *Community Energy Savings Programme*) highlighted the importance of regulatory and legislative support, cost-effectiveness (for all parties concerned) and clearly defined roles for relevant parties (for example, Lees, 2006; Lees, 2008; Eyre *et al*., 2009; UK DECC, 2011a; UK DECC, 2011b). In the USA, similar factors are apparent for state-level EERS policies (for example, Sciortino *et al*., 2011; Neubauer *et al*., 2013; Taylor *et al*., 2012). An important factor that is not only necessary for determining the success of UOs but also an inherent part of ensuring the success of future UOs is policy evaluation. As discussed in chapter two, a combination of ex-post monitoring and ex-ante estimations that comprehensively covers as much of the factors listed in figure 12 (where resources allow) is crucial to future policy success. It is clear from the documents that European evaluations tend to put greater emphasis on ex-ante approaches whereas US evaluations tend to put more weighting on ex-post approaches (Stern and Vantzis, 2014).

Labelling

Labelling (LB) refers to policies that seek to improve the communication and education of a product's energy efficiency performance. Evaluations of LB primarily focus on appliances, equipment, and buildings in the sample. The information included on labels can vary, but the evaluations concentrated mainly on energy bill savings and carbon savings. From the combined analysis, information infrastructure was the only success factor to pass the threshold, and technical issues was the key failure factor.

For success factors, the findings reveal that label design coupled with engaging awareness campaigns is crucial (Smith and Thorne, 2003; Atanasiu and

Constantinescu, 2011). As Smith and Thorne (2003) show in the evaluation of the ENERGY STAR and EnergyGuide labels in the USA, and Zheng *et al*. (2012) show in the evaluation of the Chinese equipment labelling schemes, label design has a reasonably limited impact on consumer perception of appliance quality or value if it is not coupled with engaging consumer awareness campaigns of the labelling schemes. Nadel *et al*. (2013) conducted an in-depth evaluation of the EnergyGuide labelling scheme ten years after Smith and Thorne (2003) and came to the same conclusion. Nadel *et al*. (2013) found that improvements to label design could be made by moving from a continuous-style graphic to a stars-based categorical comparison.

For failure factors, Smith and Thorne (2003) (USA), Zheng *et al*. (2012) (China), and Atanasiu and Constantinescu (2011) (EU) show that technical issues in label design and communication have impacted the success of labelling policies. The USA has had a long history of energy labelling that dates back to the *Energy Policy and Conservation Act of 1975* when the mandatory EnergyGuide label was introduced for major appliances (such as refrigerators, washing machines, tumble dryers, dishwashers and air conditioners), equipment and lighting. Smith and Thorne (2003) find that, despite a reasonable familiarity of consumers with the EnergyGuide label, it appears to have had limited impact on their product choices. The same conclusion was reached by Nadal *et al*. (2013). Zheng *et al*. (2012) found that in China there was a lack of awareness of labelling enforcement due to a lack of engagement through an initial publicity campaign.

In the EU, the *Energy Labelling of Products Directive* (Directive 2010/30/EU) was implemented by member states in 2011 (replacing the previous Directive 92/75/EC energy consumption labelling scheme) to label appliances with an energy class (colour-coded letter grade: A+++, A++, A+, A, B, C, D, E, F or G), consumption and efficiency information, noise information and general appliance details. Key appliances included in the Directive are: refrigerators, washing machines, tumble dryers, dishwashers, ovens, water heaters, hot water storage tanks, air conditioners, light bulbs, televisions, cars and tyres. A related example is the EU's *Energy Performance of Buildings Directive* (2002/91/EC), which requires member states to label buildings with Energy

Performance Certificates. Atanasiu and Constantinescu (2011) evaluated the Energy Performance Certificates and came to similar conclusions to Zheng *et al*. (2012), finding that the design of information and its communication were important issues affecting policy success. Thus, in summary, label design and communication (technical issues and information infrastructure) appear to be the key factors that transcend different countries and contexts.

Performance Standards

Performance standards (PS) refer to policies that seek to improve the energy efficiency of appliances, equipment and buildings in the manufacturing and construction process. From the combined analysis, regulatory frameworks and legislative support were the key success factors to pass the threshold, and no failure factors passed the threshold. However, failure factors were produced in the individual frequency and weighting analyses, as summarised in figure 36.

For success factors, the findings highlight that regulatory support factors are crucial. In China, the success of the building codes that were introduced in 2006 is primarily due to strong regulatory frameworks and legislative support, which includes annual governmental inspections to enforce the codes (Bin and Jun, 2012). Pengcheng *et al*. (2012) came to a similar conclusion with performance standards for appliances and equipment introduced since 1989 in China (though the evaluation period is 2009-2011). In the USA, appliance and equipment standards that were introduced since 1987 have created 340,000 jobs and resulted in US $34 billion in energy bill savings (Gold *et al*., 2011). These impacts have primarily been due to a strong regulatory drive from the national government through the *National Appliance Energy Conservation Act* (NAECA) in 1987 (Gold *et al*., 2011). As highlighted in section 5.3, performance standards have been one of the most successful categories of DSM policy.

Loans and Subsidies

Loans and subsidies (L&S) refer to policies that provide direct subsidies (grants and rebates) or indirect subsidies (no- or low-interest loans) for DSM technologies, such as energy efficiency measures or on-site generation. From the combined analysis, no success factors passed the threshold and no failure factors passed the threshold. However, success and failure factors were produced in the individual frequency and weighting analyses, as summarised in figures 35 and 36 respectively.

As the evidence base is strong for L&S policies, a possible explanation for the finding is that there is strong diversity between countries/states in terms of policy design and implementation. Furthermore, the L&S policy category contains a broad array of specific DSM policies (from tax incentives to low-interest loans, as shown in table 9 in chapter four). As such, further research should break down the results at the specific DSM policy level in order to identify if there are success and failure factors that are both frequent and highly weighted at this level.

Utility Business Models

Utility business models (UBM) refer to alternative policy and business models for energy utilities. These may take the form of decoupling policies (such as cost-recovery mechanisms) or resource standards (such as performance targets) in order to put demand-side options on an equal basis with supply-side options. Figure 10 in chapter two summarised the main UBM policies. From the combined analysis, a similar result is produced to L&S where no success factors passed the threshold and no failure factors passed the threshold. However, success and failure factors were produced in the individual frequency and weighting analyses, as summarised in figures 35 and 36 respectively.

As with L&S, the evidence base is strong for UBM policies and a similar explanation is plausible – there is strong diversity between countries/states in terms of policy design and implementation and the UBM policy category contains a broad array of specific DSM policies (from decoupling policies to system benefits charges, as shown in table 9 in chapter four). As such, further research should break down the results at the specific DSM policy level in order to identify if there are success and failure factors that are both frequent and highly weighted at this level.

Research and Development Programmes

Research and development programmes (R&D) refer to policies that aim to fund programmes that improve the research base for particular DSM technologies or help to develop and deploy them at a larger scale. This can involve small- or large-scale pilot testing, though the systematic review only concentrated on large-scale government-funded programmes, as justified in chapter three. From the combined analysis, regulatory frameworks and legislative support were the key success factors to pass the threshold, and technical issues and inadequate consumer incentives were the key failure factors.

For success factors, the findings portray that consistent governmental support over time through regulation and legislation is crucial to the success of large-scale R&D programmes. For example, in Japan, R&D programmes such as the *Sunshine Program* (introduced in 1974), the *Moonlight Program* (introduced in 1978), the *New Sunshine Program* (1993-2002), the *Super Heat Pump and Energy Accumulation Project* (1984-1992) and the *Eco Energy City Network Project* (1993-2000) were successful only where there was a long lead time for R&D investments to create outputs (Kimura, 2009). Zhou *et al*. (2011) reached similar conclusions for the Chinese on-site renewables demonstration projects in 2003 and 2009 (specifically for solar photovoltaics (PV)). In the USA, governmental support for industrial R&D programmes has been consistent since 2005 through various national policies such as EPAct 2005 (*Energy Policy Act* of 2005), EISA 2007 (*Energy Independence and Security Act* of 2007), and ARRA 2009 (*American Recovery and Reinvestment Act* of 2009) (Doris *et al*., 2009). For example, research funding efforts have focussed on improving the efficiency of energy conversion and utilisation, energy-intensive and high-emitting CO_2 processes, resource recovery and utilisation, high-temperature superconductivity wires, transmission, distribution and storage (Doris *et al*., 2009).

For failure factors, the findings convey that R&D programmes are more likely to develop technological performance and design problems and have long payback periods when they are first tested than those that are commercially available and established in markets (Kimura, 2009). As such, these technical issues lead to challenges in adequately incentivising consumers to engage with DSM technologies that are new and less established in the market.

Information Campaigns

Information campaigns (IC) refer to policies that stimulate the provision of education, training, auditing services and general marketing campaigns to relevant parties. From the combined analysis, a similar result is produced to L&S and UBM where no success factors passed the threshold and no failure factors passed the threshold. However, success and failure factors were produced in the individual frequency and weighting analyses, as summarised in figures 35 and 36 respectively.

Similar conclusions to L&S and UBM can be made for IC, as the evidence base is similarly large for IC policies compared with other categories of DSM (such as IR or R&D). Thus, there is likely to be strong diversity between countries/states in terms of policy design and implementation, as well as a broad array of specific IC policies that could be implemented (from consumer awareness campaigns to training programmes, as shown in table 9 in chapter four). As such, further research should break down the results at the specific DSM policy level in order to identify if there are success and failure factors that are both frequent and highly weighted at this level.

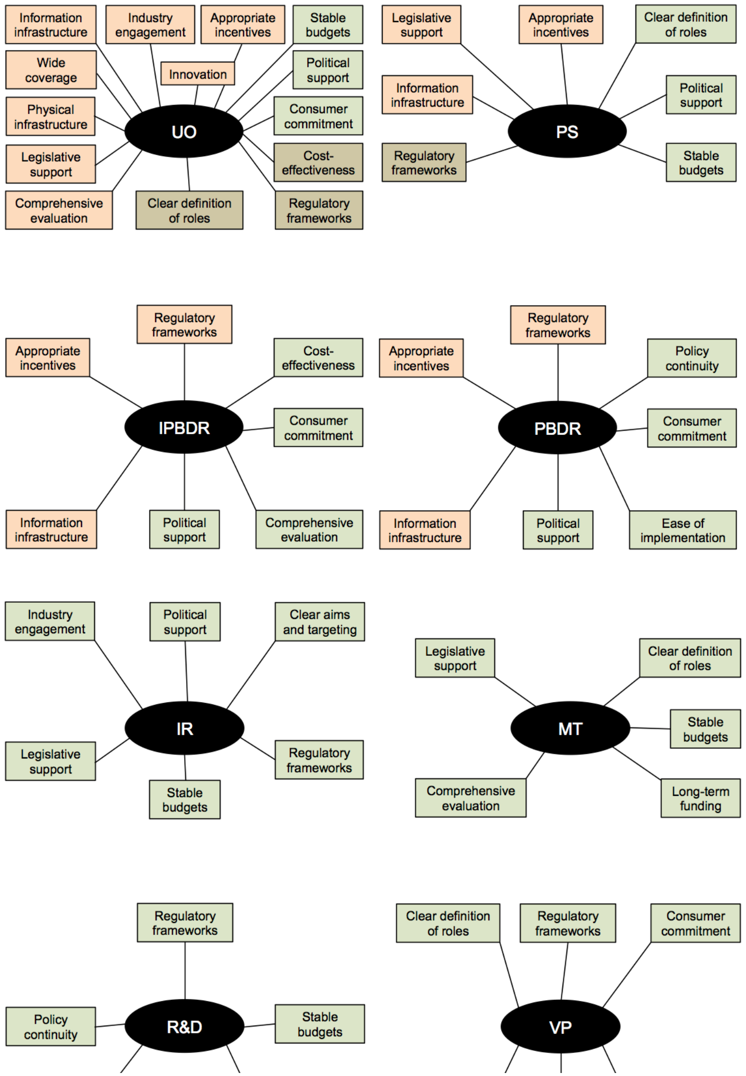
Voluntary Programmes

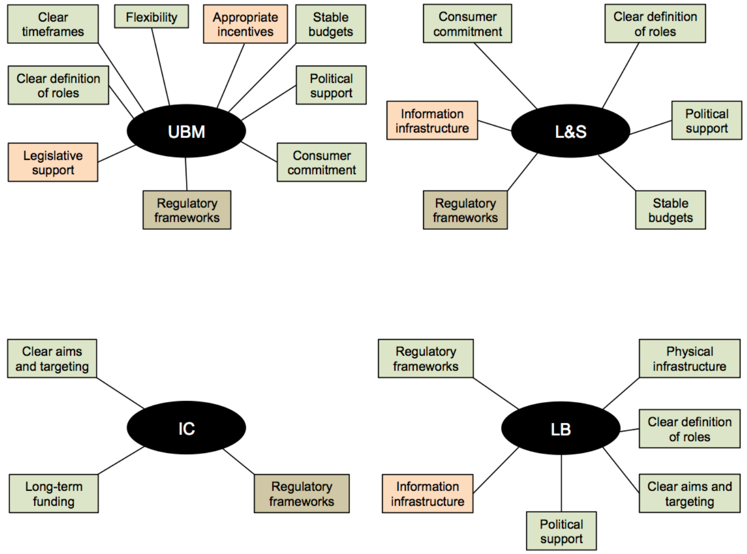
Voluntary programmes (VP) refer to all non-mandatory policies that do not fit into the other categories shown in figure 9 in chapter two. However, this category involves a number of policies that overlap between categories, such as voluntary agreements between grid operators and industrial or large commercial organisations to reduce energy consumption during peak times (which overlaps with IPBDR). From the combined analysis, clear timeframes was the only success factor to pass the threshold and no failure factors passed the threshold. However, failure factors were produced in the individual frequency and weighting analyses, as summarised in figure 36.

For success factors, the findings highlight that clear timeframes is the most crucial factor. For example, in Europe, the existence of clear goals and timeframes were important for a range of energy efficiency policies (including VP) in an evaluation conducted by Harmelink *et al*. (2008).

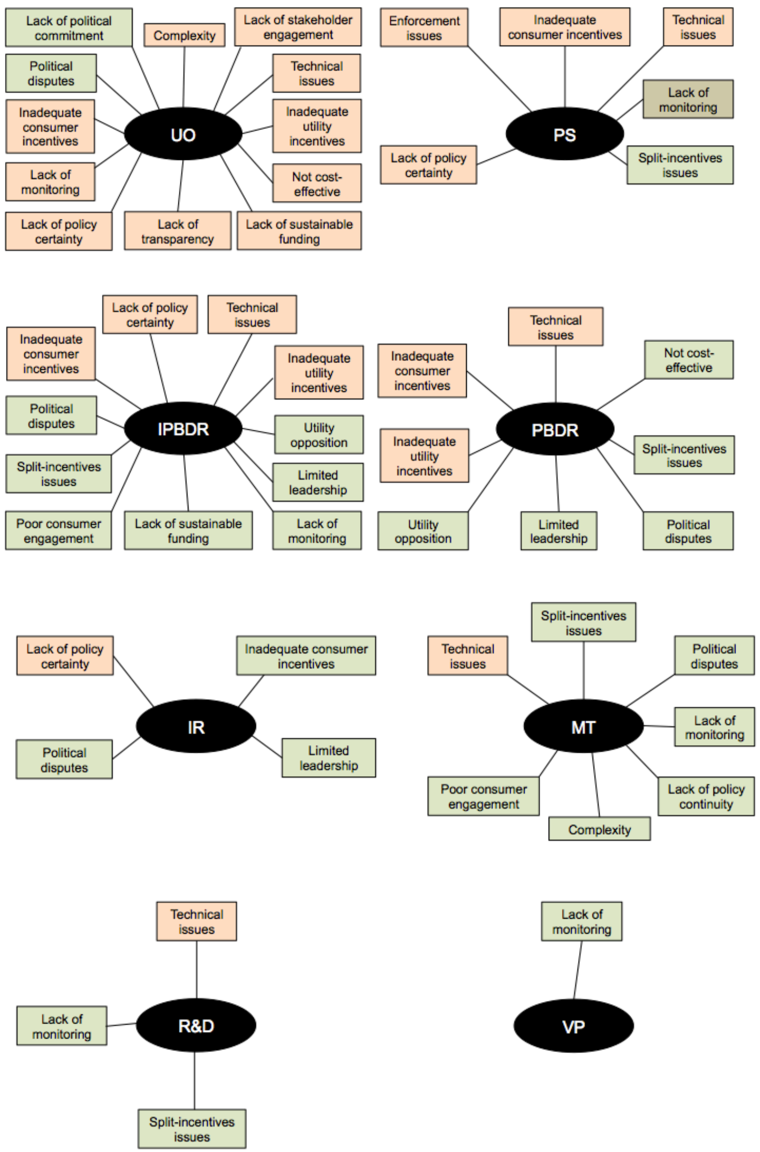
The discussions of the twelve DSM policies above are summarised in figures 35-38. Figure 35 shows the results of the individual frequency and weighting analyses for success factors by policy (excluding policy packages), and figure 36 has the same design but shows the results of the individual frequency and weighting analyses for failure factors by policy. The results for both analyses are shown together to allow comparisons to be made, though it is important to note that the figure does not show the combined analysis results, only the separate frequency and weighting results in the same diagrams. The figures are colour-coded to show the key factors that came out of each analysis: the light orange boxes are the factors from the frequency analysis, the light green boxes are the factors from the weighting analysis and the light brown boxes are the factors that overlap between analyses (though before the combined analysis equations are applied).

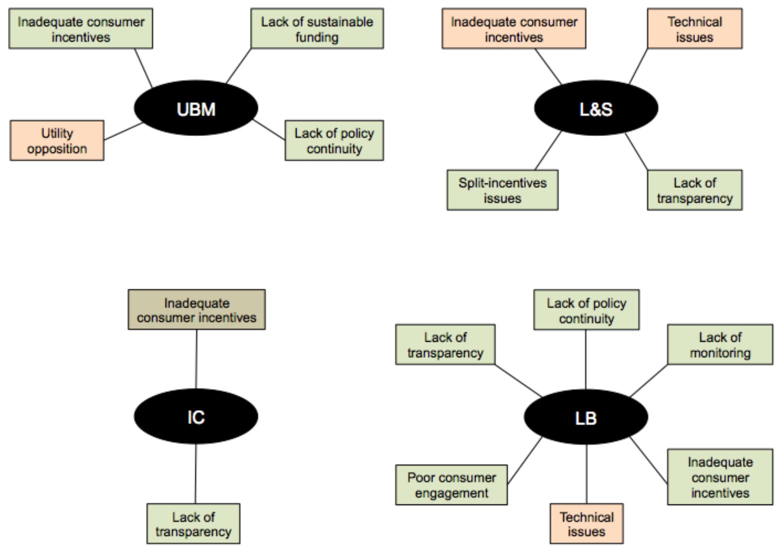
Figure 37 summarises the main results discussed in this sub-section for success factors by policy from the combined analysis. The figure has the same structure as figures 35 and 36 except that it uses a grey scale to highlight those factors that are ‘crucial’ (≥10.0% combined analysis score) and those that are ‘important’ (5.0-9.9% combined analysis score). The darker shaded boxes represent the ‘crucial’ factors and the lighter shaded boxes represent the ‘important’ factors. Unimportant factors are not included. For DSM policies where no success factors are considered crucial or important, these are grouped together. Figure 38 shows the results in the same format for failure factors from the combined analysis.





DSM Policy	Explanation
UO	Utility Obligations
PS	Performance Standards
IPBDR	Incentive Payment-Based Demand Response
PBDR	Price-Based Demand Response
UBM	Utility Business Models
L&S	Loans and Subsidies
IC	Information Campaigns
LB	Labelling
IR	Infrastructure Rollouts
MT	Market Transformations
R&D	Research and Development Programmes
VP	Voluntary Programmes
Key	**Analysis type**
Light orange boxes	Factor in frequency analysis





DSM Policy	**Explanation**
UO	Utility Obligations
PS	Performance Standards
IPBDR	Incentive Payment-Based Demand Response
PBDR	Price-Based Demand Response
UBM	Utility Business Models
L&S	Loans and Subsidies
IC	Information Campaigns
LB	Labelling
IR	Infrastructure Rollouts
MT	Market Transformations
R&D	Research and Development Programmes
VP	Voluntary Programmes
Key	**Analysis type**
Light orange boxes	Factor in frequency analysis
Light green boxes	Factor in weighting analysis
Light brown boxes	Factor in both analyses

Figure 36: the key failure factors by DSM policy from the individual frequency and weighting analyses

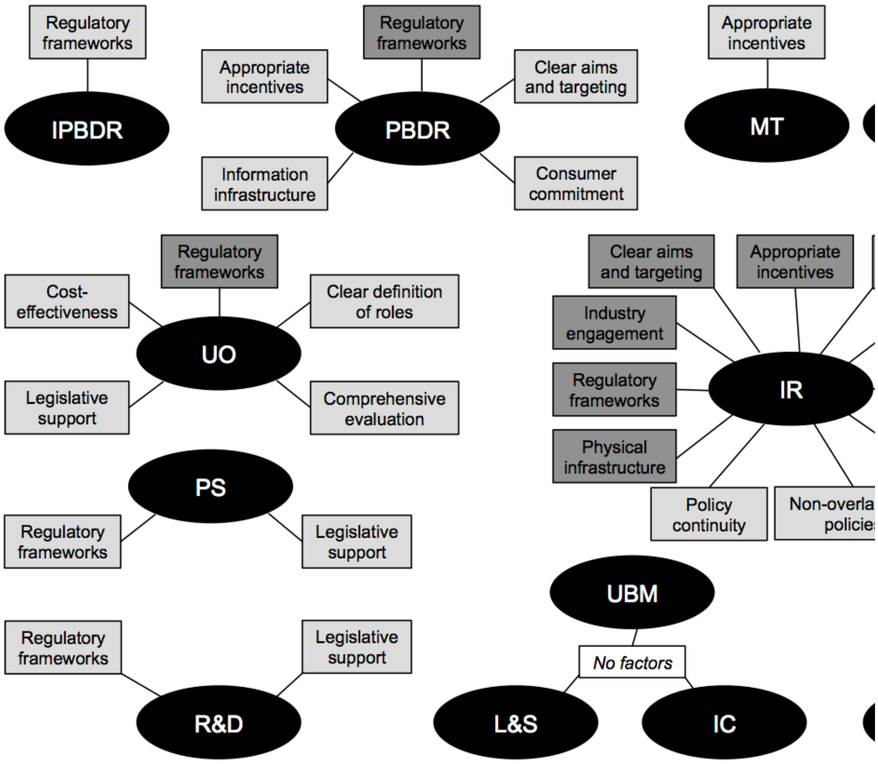


Figure 37: the key success factors by policy from the combi
results

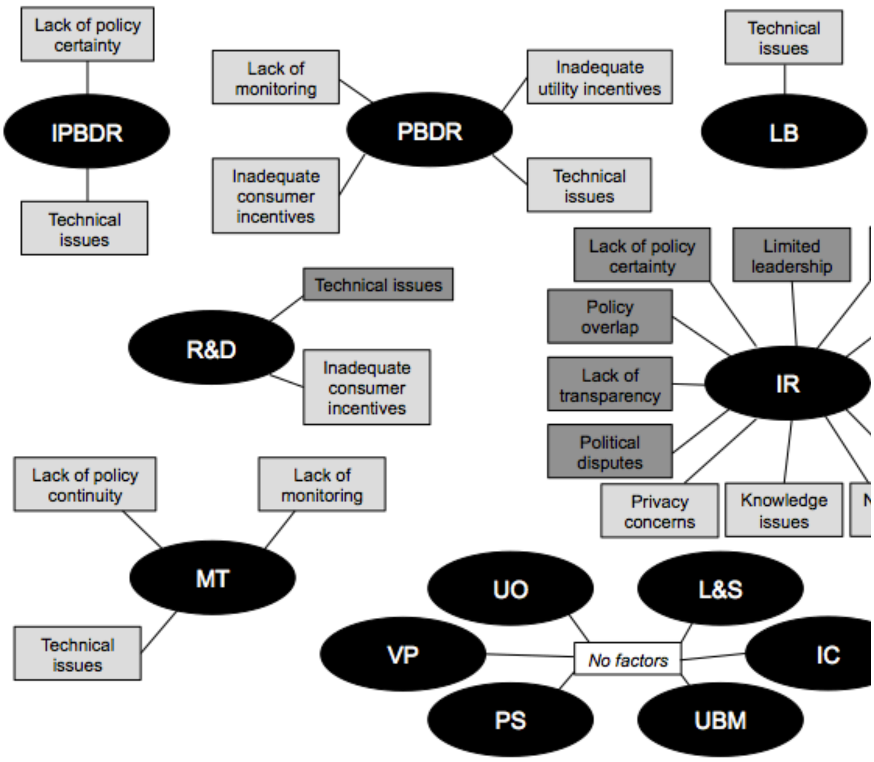


Figure 38: the key failure factors by policy from the combir
results

Table 14 summarises the top success and failure factors overall in terms of the number of policies where they are 'crucial' or 'important' factors from the combined analysis.

Success Factors	Number of Policies	Failure Factors	Number of Policies
Regulatory frameworks	6	Technical issues	6
Legislative support	4	Lack of policy certainty	3
Appropriate incentives	3	Lack of monitoring	2
Information infrastructure	3	Inadequate utility incentives	2
Clear aims and targeting	2	Inadequate consumer incentives	2
Consumer commitment	2	/	/

Table 14: the key success and failure factors overall

As table 14 shows, the top success factors across countries/states are: regulatory frameworks, legislative support, appropriate incentives, information infrastructure, clear aims and targeting and consumer commitment. Of these, regulatory frameworks is the most important success factor, as it appeared as one of the most important factors in half of the policies under examination (6/12 policies). The top failure factors across countries/states are: technical issues, a lack of policy certainty, a lack of monitoring, inadequate utility incentives and inadequate consumer incentives. Of these, technical issues (primarily programme management and administration issues) is one of the most important failure factors, as it appeared as the most important factor in half of the policies under examination (6/12 policies).

5.2.6 Key Success and Failure Factors by Country/State

This section details the key success and failure factors by country/state. As justified in chapter three, both countries and sub-national states were included in the systematic review, as some state and provincial governments have implemented and evaluated DSM policies independent of national government policy. 66 countries and states (30 countries and 36 states and provinces) were included, which were determined inductively from the sample (i.e. they were not pre-defined). The discussions group the countries/states by continental region, as shown overleaf.

- North America
- Europe
- Asia
- Australasia, South America and Africa

Due to the limited number of evaluations looking at the continents of Australasia, South America and Africa, these were included together as one continental group. The same analytical process for the frequency, weighting and combined analyses used for the examination of success and failure factors by DSM policy are used here for the examination of success and failure factors by country/state. In the previous sub-section, the results for each DSM policy were averaged across all countries/states, whereas in this sub-section, the results for each country/state are averaged across all DSM policies. A small difference in the combined analysis equations is that the word 'policy' is replaced with 'country/state' (e.g. 'policy success weighting' becomes 'country/state success weighting') though the calculations remain the same. Nevertheless, a worked example is shown below for one country/state to show how the results were derived for each of the 66 countries/states in the sample.

<u>*Importance of Comprehensive Evaluation for the UK:*</u>
<u>Success Factor:</u> Comprehensive Evaluation (EV)
<u>Success Factor Frequency:</u> 12 (above ≥5 frequency threshold)
<u>Success Factor Weighting:</u> 2.8 (in 2.5-3.0 'crucial' weighting group)
<u>Country/State Frequency:</u> 22 (number of evaluations focussing on the UK in the total sample of 690 evaluations)
<u>Country/State Weighting:</u> 3.8 (averaged weighting of country/state across the evaluations)
<u>Factor Combined Analysis Score:</u> 3.8 x (12 x 2.8) / 10 = 12.9
<u>Country/State Theoretical Maximum Score:</u> 3.8 x (22 x 5) / 10 = 41.8
<u>Combined Analysis Percentage Score:</u> (12.9 / 41.8) x 100% = 30.9%

The example produces a score of 30.9%, which strongly fits into the 'crucial' category of the combined analysis. Thus, comprehensive evaluation is one of the most important factors determining the success of DSM policies in the UK.

Within each continental grouping, case studies of specific countries/states are given. However, the full results for each of the 66 countries/states for each of the three analysis types (frequency, weighting and combined) are summarised in figures 39-42 at the end of the sub-section.

North America

From the combined analysis, fourteen North American countries/states produced success factors (USA, USA (state-level), Canada, Mexico, the US states of California, New York, Vermont, Wisconsin, Ohio, Texas and Florida, the US region of PJM, and the Canadian provinces of British Columbia and Ontario). The top success factors are:

- Regulatory frameworks (for twelve countries/states)
- Comprehensive evaluation (for nine countries/states)
- Innovation (for nine countries/states)
- Appropriate incentives (for seven countries/states)
- Policy continuity (for seven countries/states)
- Industry engagement (for seven countries/states)
- Consumer commitment (for seven countries/states)

Thirteen North American countries/states produced failure factors in the combined analysis (USA, Canada, the US states of California, New York, Vermont, Massachusetts, Maine, Connecticut, Ohio, Florida, Texas, Wisconsin and the US region of the Pacific Northwest). The top failure factors are:

- A lack of monitoring (for nine countries/states)
- Technical issues (for eight countries/states)
- Inadequate utility incentives (for six countries/states)
- A lack of policy certainty (for five countries/states)
- A lack of sustainable funding (for five countries/states)

It is clear from the North American sample that the full range of success factor categories is needed for DSM policies to be successful (regulatory support, financial support, policy characteristics, stakeholder engagement and infrastructure, as per the categories in figure 30). For example, regulatory frameworks and consumer commitment were crucial for Mexico’s Compact Fluorescent Lighting (CFLs) subsidy programme from 1995-1997, as shown in an evaluation conducted by the International Institute for Energy Conservation (IIEC). The evaluation describes how the Mexican government mandated the Comisión Federal de Electricidad (CFE) to purchase CFLs in bulk under competitive procurement from manufacturers (and receive a significant discount over the retail market price) and then sell the CFLs to consumers at a subsidised price (60% cheaper than the market price). The evaluation highlights how extensive consumer marketing and outreach were undertaken to encourage consumers to engage with the subsidy.

In a further example of subsidy policies, an evaluation by the IEA DSM Programme (2004) conveyed the importance of appropriate rebates for industrial energy savings opportunities in Canada. The evaluation describes how proposed projects with savings of <200 MWh/year and payback periods of two years received up to CAN $0.10 for first-year kWh savings. The rebates appear to have worked well, as they provided “just enough” incentive to motivate customer implementation (IEA DSM Programme, 2004).

The findings highlight that regulatory support and appropriately incentivising industry are the most crucial factors for the USA. For certain policies, such as demand response (IPBDR and PBDR), information on the actual performance of demand response resources is also central to assessing the long-term viability of such resources, as evaluations such as Cappers *et al*. (2009) have shown. Goldman *et al*. (2011) highlighted how ARRA (*American Recovery and Reinvestment Act of 2009*)-funded programmes devoted significant financial support to energy efficiency (as separate from utility customer-funded programmes), which led to rapid job creation and retention as well as a reduction in energy use and greenhouse gas emissions (though the evaluation did not quantify these impacts). Nevertheless, it appears that regulatory support is the most crucial factor category and the majority of the evaluations discuss factors within this category (such as regulatory frameworks and legislative support). For example, Doris *et al*. (2009) argue that strong leadership through a central administrative body, such as the Environmental Protection Agency (EPA), is necessary for developing the rules and implementing similar DSM policies in different states. The evaluation uses the example of building codes to highlight the importance of legal procedures for enforcement.

Rigorous measurement, verification and open evaluation are also central to the improvement of future policies (Bachrach, 2003; Kushler and Vine; Zuckerman *et al*., 2013; O'Drain and Edwards, 2010). Evaluations should involve ex-post monitoring and how comprehensive the monitoring is depends on the resources available, though DiSanto (2014) argues that at least 1% of DSM programme budgets should be dedicated to evaluation. O'Drain and Edwards (2010) emphasise the crucial need for close interaction between utilities and commissioners to improve policies in subsequent policy periods. The evaluation looks at how the 2012-2014 policy period for the LIEE (Low Income Energy Efficiency) programme could be improved through better engagement with industry to ensure enhanced dialogue and agreement between four Investor-Owned Utilities (IOUs) and the CPUC (California Public Utilities Commission) in California.

Despite these successes, the examples of failure in the North American sample are equally interesting. For example, Bachrach (2003) identifies technical issues as a key challenge in terms of who should administer energy efficiency programmes. Furthermore, technical issues occur where conflicts arise between maximising cost-effective energy savings and trying to meet equity objectives, and coordination and practical problems are evident (Kushler and Vine, 2003). This can lead to inadequate incentives for utilities to undertake detailed programme monitoring, particularly where incentives and penalties are based on performance. For example, Zuckerman *et al*. (2013) show how the introduction of the *Risk/Reward Incentive Mechanism* (RRIM) in 2007 in California (a shareholder incentive for utilities to deliver energy efficiency programmes) created difficulties in the evaluation of energy savings, which led to conflicts between utilities and the CPUC. The first phase of the policy (2007-2009) included an ex-post 'true-up' to identify whether or not utilities had met their targets in the determination of fines or rewards. Utilities opposed this, as they could not then adapt their programmes to meet targets if they had been underperforming and the resulting disputes led to the removal of the ex-post 'true-up' in the second phase (2010-2012) (Zuckerman *et al*., 2013).

In the USA, the decentralised nature of political power in the country, where individual states have more regulatory control than their equivalents in other countries (such as Chinese provinces or European local governments), forms a strong part of the reason for the challenges in reaching agreement for implementing national-level DSM policies. The source of sustained, non-fluctuating funding and incentives (Goldman *et al*., 2011), and the degree of policy certainty in the delivery of sustainable funding, are crucial at the national level in the USA. However, since 2005, the country has been active in overcoming this factor with DSM funding provided through the *Energy Policy Act* (EPAct) (2005), the *Energy Independence and Security Act* (EISA) (2007), ARRA (2009), and the *Energy Efficiency Improvement Act* (EEIA) (2015) (at the time of writing this bill has not yet been enacted and is currently being considered by the Senate after passing the House of Representatives).

Europe

From the combined analysis, twelve European countries/states produced success factors (Italy, Flanders (Belgium), France, Denmark, Spain, Germany, the UK, the EU, Sweden, Netherlands, Ireland and Croatia). The top success factors are:

- Regulatory frameworks (for eleven countries/states)
- Appropriate incentives (for eleven countries/states)
- Comprehensive evaluation (for ten countries/states)
- Legislative support (for eight countries/states)
- Industry engagement (for seven countries/states)
- Innovation (for seven countries/states)

Eleven European countries/states produced failure factors from the combined analysis (Germany, Italy, Denmark, the UK, Flanders (Belgium), Belgium, Netherlands, Spain, France, Sweden and the EU). The top failure factors are shown overleaf.

- A lack of monitoring (for eleven countries/states)
- A lack of sustainable funding (for seven countries/states)
- Technical issues (for seven countries/states)
- Complexity (for six countries/states)
- Inadequate utility incentives (for six countries/states)

It is clear from the European sample that, like the North American sample, the full range of success factor categories is needed for DSM policies to be successful (regulatory support, financial support, policy characteristics, stakeholder engagement and infrastructure, as per the categories in figure 30). The utility obligations in the UK, Italy, France and Denmark have been successful particularly as a result of strong regulatory and legislative support, as well as industry engagement (Eyre *et al*., 2009; Bundgaard *et al*., 2013). In the UK, these factors were strong influences on developing obligations that were cost-effective and appropriately incentivised utilities to meet targets. For example, the *Energy Efficiency Commitment 1*, which ran from 2002-2005, led to suppliers exceeding their targets by 18% (Lees, 2006).

Furthermore, an evaluation of the UK's following obligation, the *Energy Efficiency Commitment 2*, which ran from 2005-2008, showed that suppliers delivered 23% more cost-effectively than the government's illustrative mix (Lees, 2008). Evaluations of the UK's utility obligations similarly highlight the importance of comprehensive evaluation and some of the challenges, such as the need to determine additionality and the impacts of behavioural changes (UK DECC, 2011c; Rosenow and Galvin, 2013; Giraudet *et al*., 2012). A large number of the evaluations in the European sample evaluate utility obligations, primarily in the UK, Italy, France and Denmark. However, in Germany, the evaluations focus on loans and subsidies and information campaigns, such as policies for energy auditing and marketing campaigns to improve the uptake of efficient appliances (for example, Fleiter *et al*., 2012; IEA DSM Programme, 2004).

For policy failure, the findings convey how budget constraints can affect the stability and consistency of funding over the long-term (Rosenow, 2011; Gruber *et al*., 2011). Furthermore, Harmelink *et al*. (2008) found that monitoring and verification of actual energy savings have a relatively low priority in a number of European countries. This is often linked to funding levels and policy priorities. As previously discussed, ex-ante approaches tend to be more frequently used in Europe than ex-post approaches, which are more common in North America (Stern and Vantzis, 2014). For example, for the utility obligations in Italy, ~90% of certified savings are determined using deemed savings (Waide and Buchner, 2008).

Bundgaard *et al*. (2013) evaluated Danish utility obligations and took into account the technical accuracy in the savings calculations, additionality, rebound effects and spill-over effects. Nevertheless, the results are still based on reported savings by utilities and due to the lack of penalties and few independent random sampling tests being carried out, there are limited incentives for utilities to report savings accurately. As such, utilities have an incentive to over-report savings (Bundgaard, 2013), as was found with the Californian RRIM.

Technical issues and complexity are crucial factors that have led to examples of policies not performing as well as anticipated in Europe. For example, DECC (2011a) identified that there were a number of technical issues in the *Community Energy Savings Programme* (CESP), which ran from 2009-2012 as an obligation on suppliers. The evaluation highlights that there needed to be a greater emphasis on professional installation standards, a broader array of delivery partners to spread risks and a reduction in the technical monitoring failure rates (which were estimated to be ~15%) (DECC, 2011a). A number of utilities did not meet their targets for CESP, such as British Gas (fined £11.1 million), Scottish Power (fined £2.4 million), Scottish and Southern Energy (fined £1.75 million), and GDF Suez (fined £450,000) (McClone, 2014c). Another example is the *Low Carbon Buildings Programme*, which ran from 2006-2011 as a (primarily) grants-based policy (DECC, 2011b). DECC (2011b) identified that the scheme was complex to administer and did not manage to incentivise significant numbers of additional installations that would not have taken place in the absence of the policy.

Asia

From the combined analysis, nine Asian countries/states produced success factors (China, Japan, India, the Philippines, Thailand, South Korea, Orissa (India) and the Chinese provinces of Hebei and Fujian). The top success factors are:

- Regulatory frameworks (for seven countries/states each)
- Industry engagement (for seven countries/states each)
- Innovation (for six countries/states)
- Ease of implementation (for five countries/states)
- Cost-effectiveness (for five countries/states)
- Policy continuity (for five countries/states)

Eighteen Asian countries/states produced failure factors (China, Japan, India, the Philippines, Thailand, South Korea, Orissa (India), Sri Lanka, Pakistan and the Chinese provinces of Beijing, Shanghai, Hefei, Guangzhou, Shandong, Sichuan, Fujian, Hebei and Jiangsu). The top failure factors are:

- Technical issues (for seventeen countries/states)
- A lack of monitoring (sixteen countries/states)
- Knowledge issues (for eleven countries/states)
- A lack of sustainable funding (for ten countries/states)
- Enforcement issues (for nine countries/states)
- Poor consumer engagement (for nine countries/states)

Yu (2010) argues that in China, regulatory support is important, as experience in other countries has shown that a competitive market does not automatically deliver energy efficiency. Vine *et al*. (2006) reiterate this point for equipment labelling policies in South Korea. In China, investment in research and development to stimulate innovation in energy efficiency and the focus on commencing initiatives from single points (places) and then expanding to wider areas has helped to ease the implementation process (Bin and Jun, 2012). These policies have primarily revolved around building codes, building labelling and demand response (Bin and Jun, 2012). Zheng *et al*. (2012) also show how stakeholder engagement was important in the success of equipment standards and labelling in China in the first policy period (2006-2007), as it helped to improve the policies in the second policy period (2009). A similar result for the Philippines is discussed in an evaluation conducted by the IIEC, where technical seminars and technology presentations were held during site visits to improve stakeholder engagement for energy efficiency subsidy programmes and demonstration programmes. Thus, regulatory support and stakeholder engagement appear to be crucial to DSM policy success in east-Asian countries.

RAP (Regulatory Assistance Project) (2012) argues that China's economic and political structure encourages long-term national planning and policy continuity, which have been key factors for a range of DSM policies, such as performance standards, loans and subsidies, information campaigns, labelling, demand response and research and development programmes. Pengcheng *et al*. (2012) also highlight the cost-effectiveness of Chinese DSM policies, particularly appliance and equipment standards.

Technical issues appears to be the dominant failure factor for the Asian countries/states in the sample. For example, documents that evaluated policies for Japan found that, despite general success with DSM policies, technical barriers in research and development programmes were important issues, especially in relation to organisational restructuring, the stagnation of related markets and the long period of time for the commercialisation of technologies (Kimura, 2009).

In China, the findings highlight that the enforcement of some DSM policies, such as performance standards for buildings and equipment, is challenging (Hong, 2009; Bin and Jun, 2012), and this ties closely with monitoring and testing issues (Zheng *et al*., 2012). Zheng *et al*. (2012) highlight issues with a lack of awareness and the absence of an initial publicity campaign on efficiency standards. These knowledge issues led to manufacturers failing to register their products, as well as retailers being resistant to inspection (Zheng *et al*., 2012). This also links strongly with inadequate incentives for manufacturers and retailers to pursue energy efficient designs and materials (Zhou *et al*., 2011).

Despite this, China has generally been successful with DSM policies, as shown in section 5.3.

A final example is India. DSM policy is not well developed in the country and due to rapid growth in energy consumption, there is increasing attention being given to demand-side activities. RAP (2012) argues that the key issues in India are a lack of robust monitoring and verification systems and enforcement mechanisms for DSM. This draws strong parallels to China, which is also facing rapid growth in energy consumption. However, per capita energy consumption is still much lower in India than in developed countries (~0.61 toe/capita compared with ~2.97 toe/capita in the UK or 3.81 toe/capita in Germany) (World Bank, 2011). This is similar to some extent with China (~2.03 toe/capita, though this is now close to that of developed countries) (World Bank, 2011).

Australasia, South America and Africa

The remaining continents: Australasia, South America and Africa (excluding Antarctica), are included together within the same grouping as the evidence base for DSM policies outside of North America, Europe and Asia is limited. From the combined analysis, eight countries/states produced success factors (Brazil, Australia, Kenya, South Africa and the Australian states of the Australian Capital Territory, New South Wales, Victoria and South Australia). The top success factors are:

- Legislative support (for seven countries/states)
- Comprehensive evaluation (for six countries/states)
- Innovation (for six countries/states)
- Regulatory frameworks (for five countries/states)
- Appropriate incentives (for five countries/states)
- Industry engagement (for five countries/states)

Nine countries/states produced failure factors (Australia, the Australian states of the Australian Capital Territory, New South Wales, South Australia and Victoria, New Zealand, South Africa, Kenya and Brazil). The top failure factors are shown overleaf.

- A lack of monitoring (for eight countries/states)
- Complexity (for seven countries/states)
- Technical issues (for seven countries/states)
- Not cost-effective (for five countries/states)
- A lack of transparency (for four countries/states)

In South America, Brazil has the greatest evidence base for DSM policy evaluation. The use of the International Performance Measurement and Verification Protocol (IPMVP) for evaluating its utility obligation has been a successful way of aligning its evaluation work with international experiences and standards (Broc *et al*., 2012). Although Lepetitgaland *et al*. (2011) argue that there are issues with the evaluation of Brazil's DSM policies generally, the document details specific ways that the evaluation process could be improved. The Brazilian evaluations are good examples of the importance of evaluation in this continental grouping. In addition, the findings highlight that regulatory support is equally crucial, which draws parallels to the importance of regulatory frameworks in North America, Europe and Asia. For example, in 2000, the *National Law 9.991/2000* established that Brazilian electric utilities should apply a minimum of 0.5% of revenues to energy efficiency programmes (Broc *et al*., 2012). This led to PEE (the *Energy Efficiency Program*), a policy introduced in 2000 and which applies to electricity distribution utilities (Lepetitgaland *et al*., 2011). ANEEL, the Brazilian Electricity Regulatory Authority, sets the rules for PEE and supervises a utility's energy efficiency projects and since 2010 (as a result of the *Bill 12.212/2010*), 60% of the investments need to be performed in low-income communities and households on social tariffs (Broc *et al*., 2012). Despite this, Brazil has had regulatory support for DSM since 1985, when its flagship policy, PROCEL (the *Federal Electricity Conservation Program*), was established (Lepetitgaland *et al*., 2011). PROCEL uses various DSM policies, such as research and development funding to stimulate innovation, education programmes targeted at end-users and low-interest financing for utilities to undertake energy efficiency projects.

Linked to regulatory frameworks is the importance of legislative support and in South Africa, it is provided through two main channels – firstly, DSM activities are managed by a national DSM department that was set up in Johannesburg and secondly, regional targets are incorporated into the regional performance measurements of the six Eskom Distribution regions where an Energy Services Manager (in each region) is responsible for achieving the targets (IEA DSM Programme, 2008). Africa is the continent with the weakest evidence base for DSM policy and outside of South Africa, Kenya is the only other African country in the sample. Although Guertler and Royston (2013) generally find that the Kenyan micro-finance programme for DSM, primarily for micro-generation (such as solar PV), has been unsuccessful, the evaluation highlights the importance of appropriate incentives and engagement with consumers and industry.

Four Australian states represent the greatest evidence base in Australasia (the Australian Capital Territory, New South Wales, Victoria and South Australia). Although Australia has been less successful with implementing and evaluating DSM policies at the national level, the findings convey the emphasis that the national government places on engaging the energy industry and relevant stakeholders. Effendi and Courvisanos (2012) argue that this is partly due to the political power of the energy and resources industry (particularly fossil fuels and mining) in the country. To some extent, the evidence base for Australia draws parallels to that of the USA, where more of the DSM policy experiences have taken place at a state-level rather than at a national level. The utility obligations in the Australian Capital Territory, New South Wales, Victoria and South Australia have dominated the evaluation evidence base in Australia.

Despite the successes in the continental grouping of Australia, South America and Africa, Broc *et al*. (2012) emphasise how in Brazil, despite the use of the IPMVP in the evaluation of its utility obligations, there were no extensive training programmes about the procedures of the IPMVP or certification for people evaluating the programmes. As there were technical difficulties in adapting the IPMVP options to the programmes, utilities often just made rapid measurements a week before and after the installation of energy efficiency measures (Broc *et al*., 2012). As ANEEL (the Brazilian electricity regulatory agency) did not conduct ex-post verifications of the utility obligations, the utilities were responsible for monitoring and evaluating their projects, thus creating a non-transparent environment where there was no independent verification of the programmes (Broc *et al*., 2012). As such, Lepetitgaland *et al*. (2011) argue that the evaluation process is not properly valued in Brazil.

In the Australian states of the Australian Capital Territory and New South Wales, technical issues in policy design and evaluation emerged as key factors for their utility obligations. For example, under a cap-and-trade scheme, in contrast to a baseline-and-credit design, double counting of abatement can occur where offsets provided for energy efficiency have been undertaken in the same sector as the obligated parties (Crossley, 2008). Complexity and technical issues in policy design were also evident in New Zealand's 'Electricity Governance Rules' for its ancillary services market, which was developed in 1996 (Crossley, 2008). A number of electricity providers invested in interruptible load-based DSM (a type of IPBDR), but due to increased compliance requirements in 2004, many providers could not meet these requirements so withdrew (or new providers did not enter the market) (Crossley, 2008).

The discussions of the four continental groupings above are summarised in figures 39-42. Figure 39 shows the key success factors for each country/state (excluding policy packages) by frequency and weighting. The countries/states are grouped together by continent. The results for both frequency and weighting are shown together to allow comparisons to be made, though it is important to note that the figure does not show the combined analysis results. Figure 40 shows the results in the same format for failure factors. The figures are colour-coded to show the key factors that came out of each analysis (frequency and weighting): the light orange boxes are the factors from the frequency analysis, the light green boxes are the factors from the weighting analysis and the light brown boxes are the factors that overlap between analyses (though before the combined analysis equations are applied).

Figures 41 and 42 are the more important visualisations, as they show the final results for success factors (figure 41) and failure factors (figure 42) from the combined analysis. The figures have the same structure as figures 39 and 40 except that they uses a grey scale to highlight those factors that are 'crucial' (≥10.0% of the theoretical maximum score) and those that are 'important' (5.0-9.9% of the theoretical maximum score). The darker shaded boxes represent the ‘crucial’ factors and the lighter shaded boxes represent the ‘important’ factors. Unimportant factors are not included. For countries/states where no success factors are considered crucial or important, these are grouped together.

Figure 39: the key success factors by country/state for both the frequency and weighting analyses

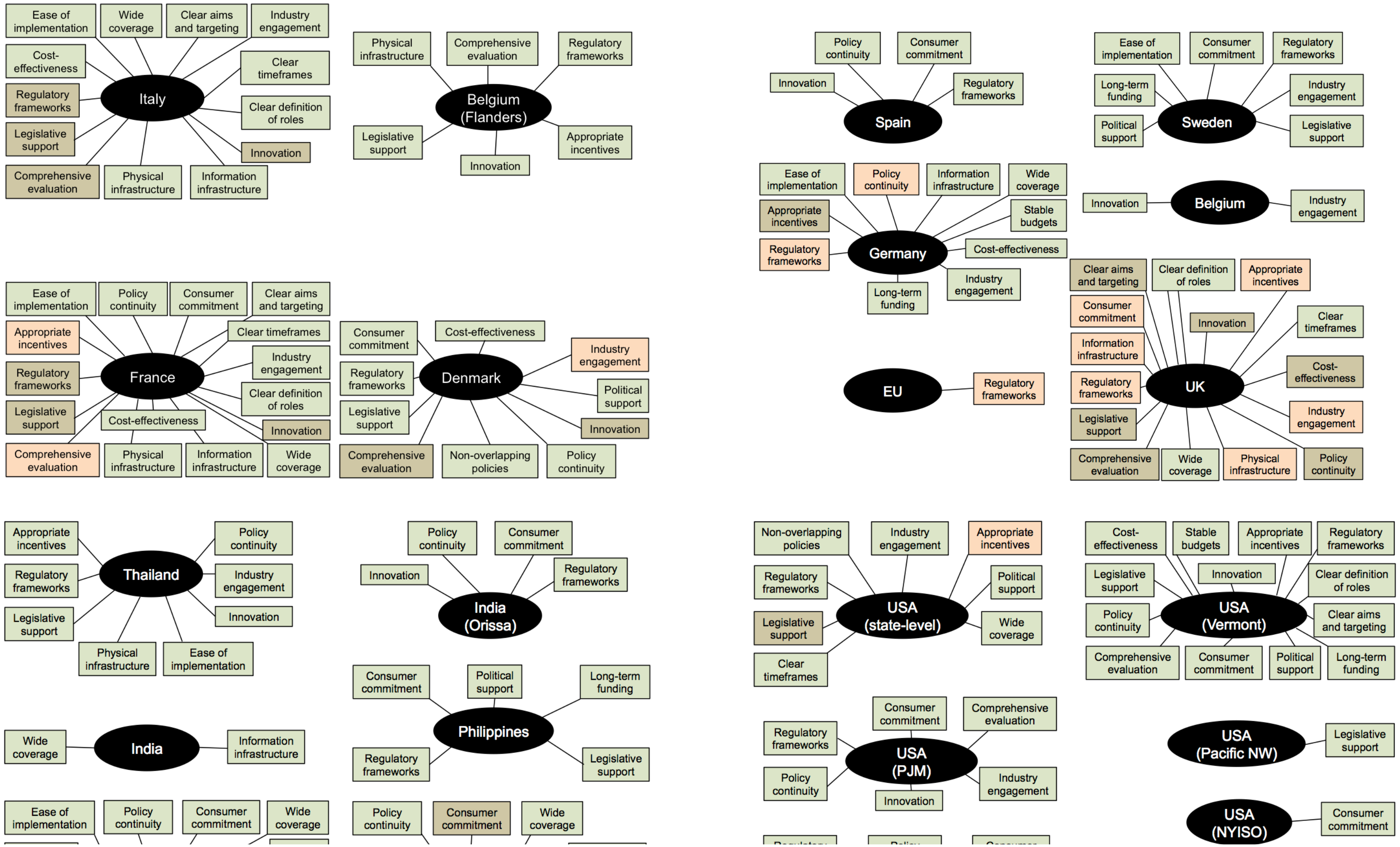
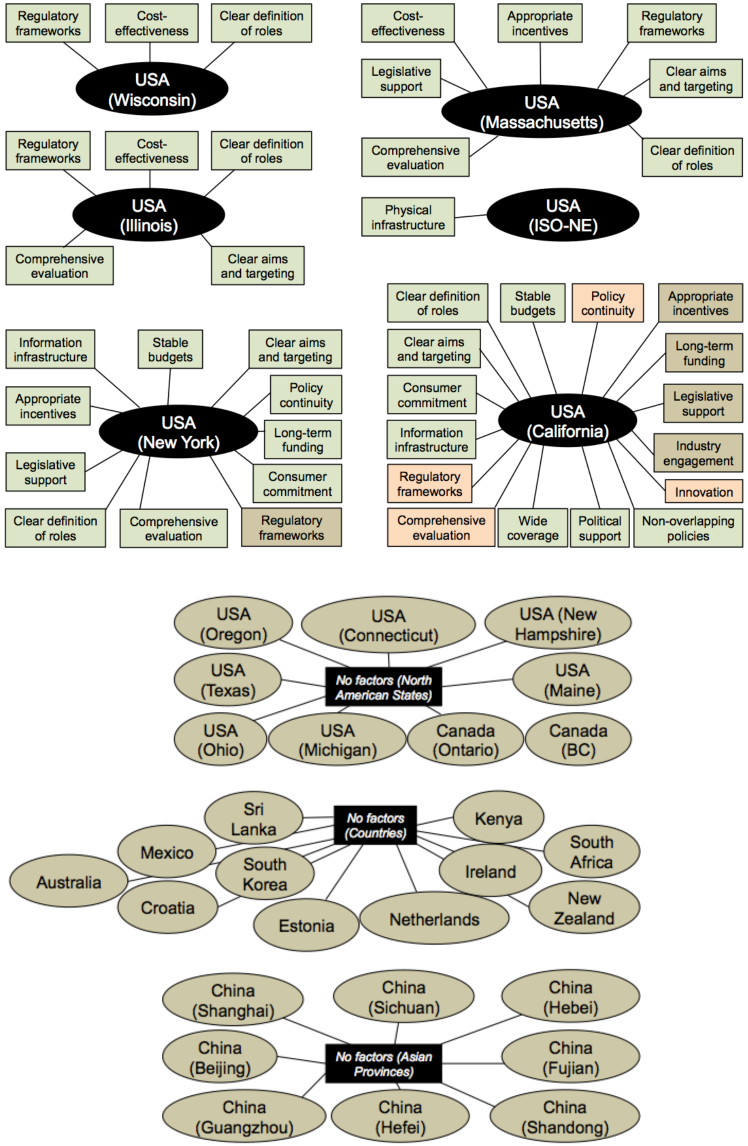
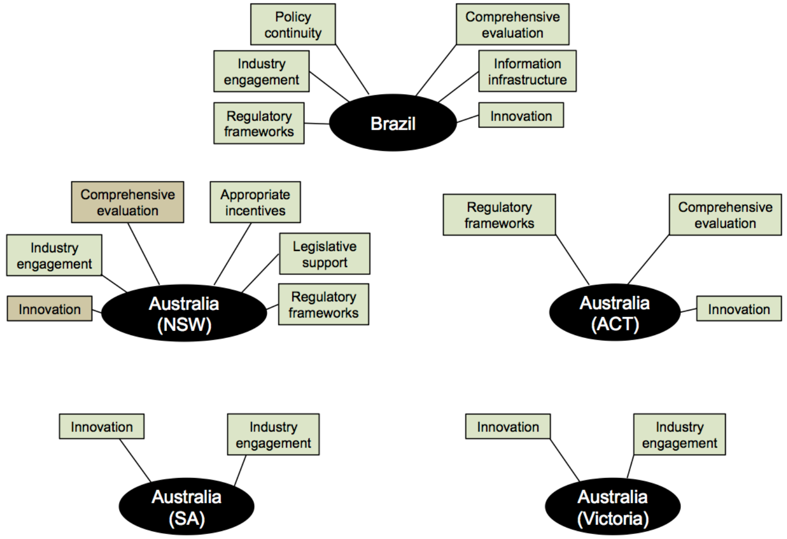


Figure 39 (continued)





Key	Analysis type
Light Orange Boxes	Factor in frequency analysis
Light Green Boxes	Factor in weighting analysis
Light Brown Boxes	Factor in both analyses

Figure 40: the key failure factors by country/state for both the frequency and weighting analyses

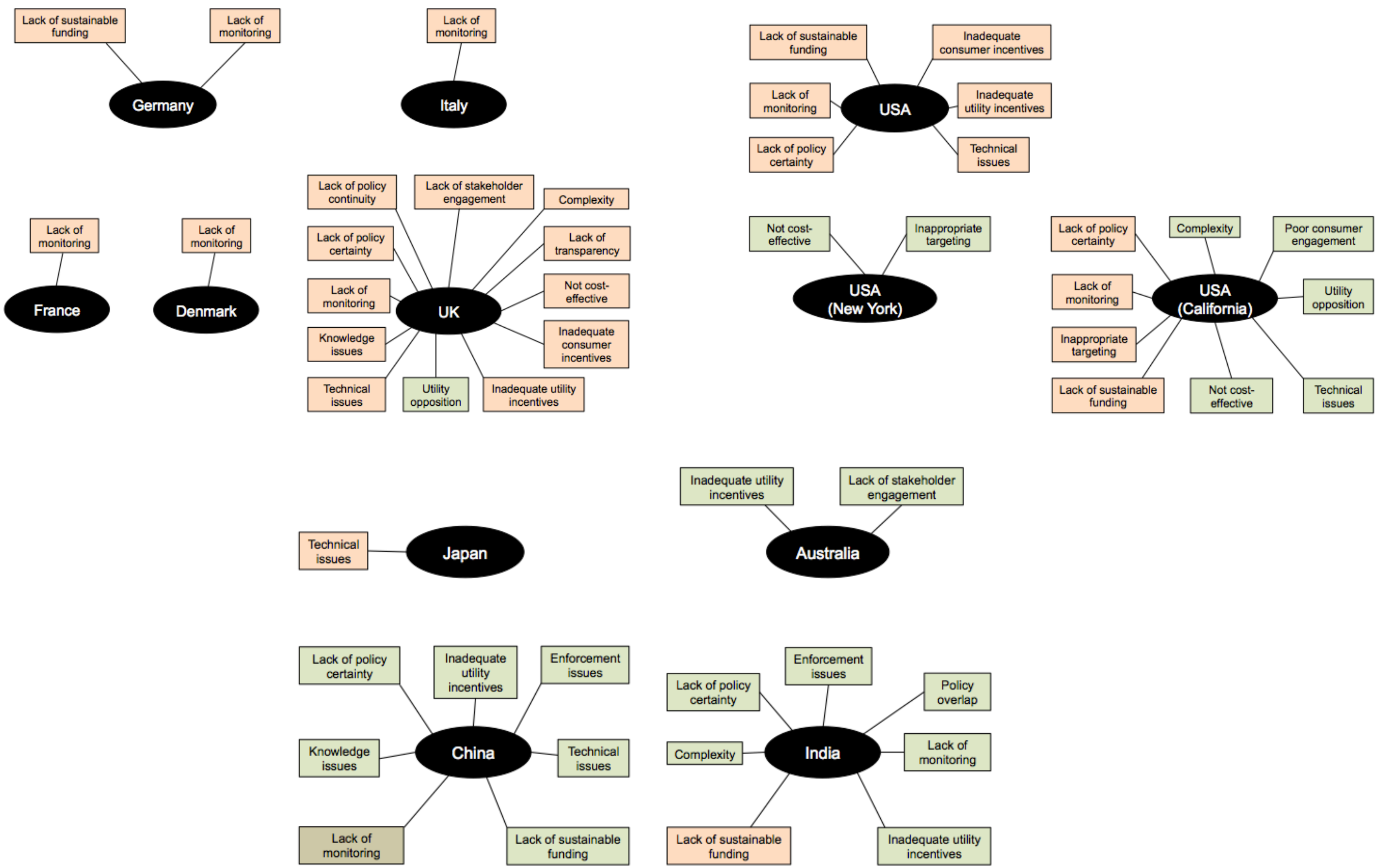
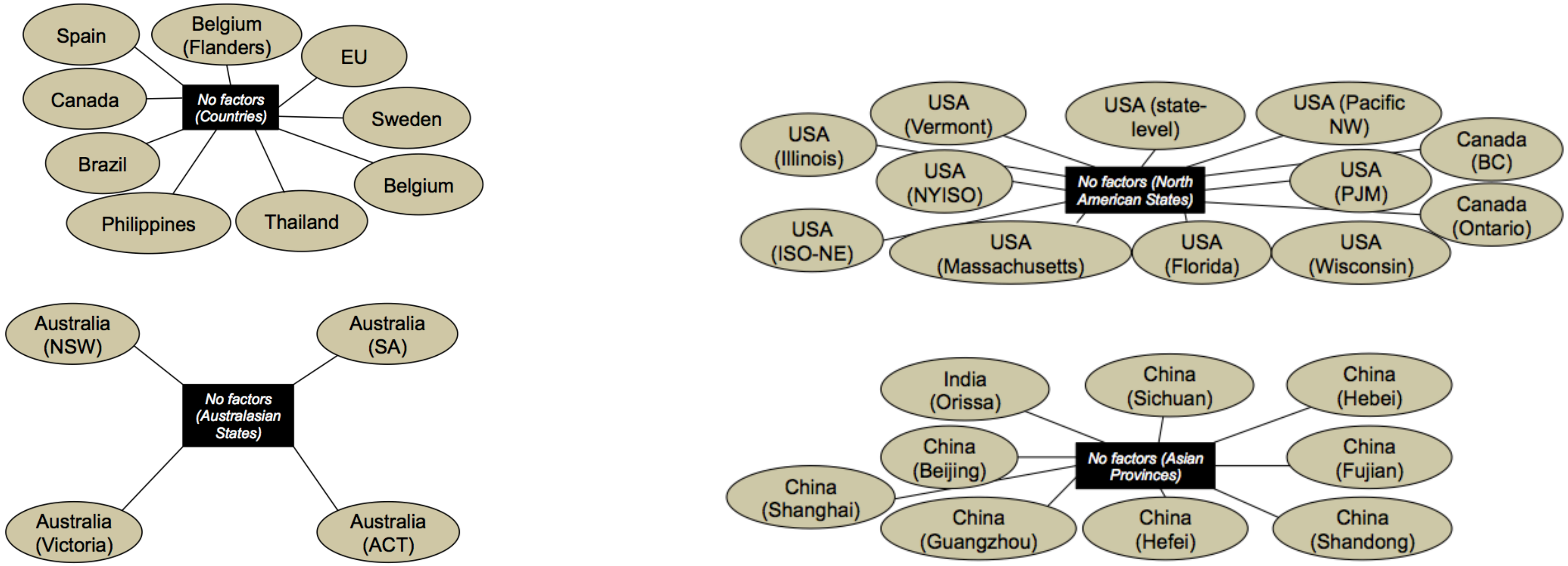


Figure 40 (continued)



Key	**Analysis type**
Light Orange Boxes	Factor in frequency analysis
Light Green Boxes	Factor in weighting analysis
Light Brown Boxes	Factor in both analyses

Figure 41: the key success factors by country/state based on the combined analysis results

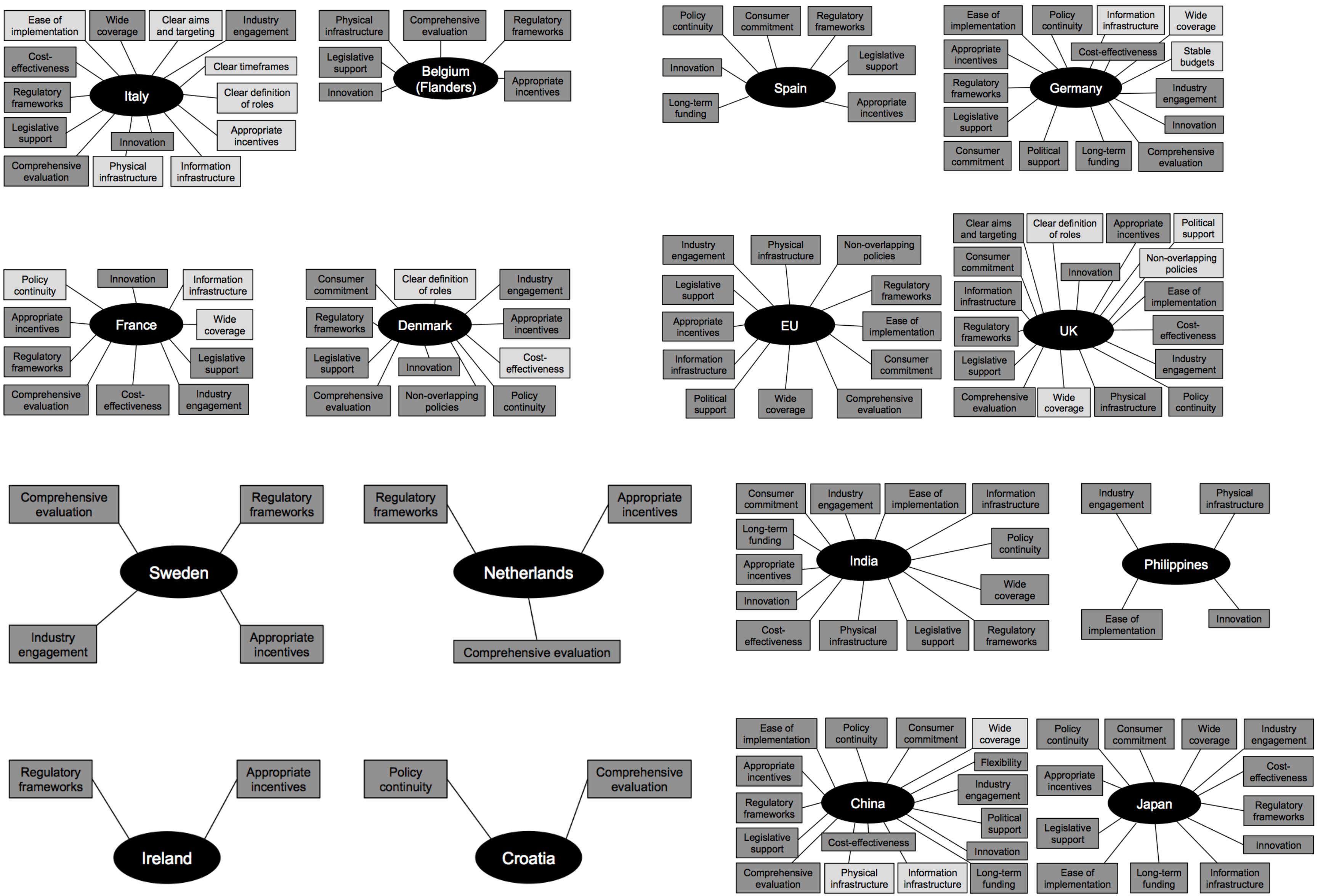


Figure 41 (continued)

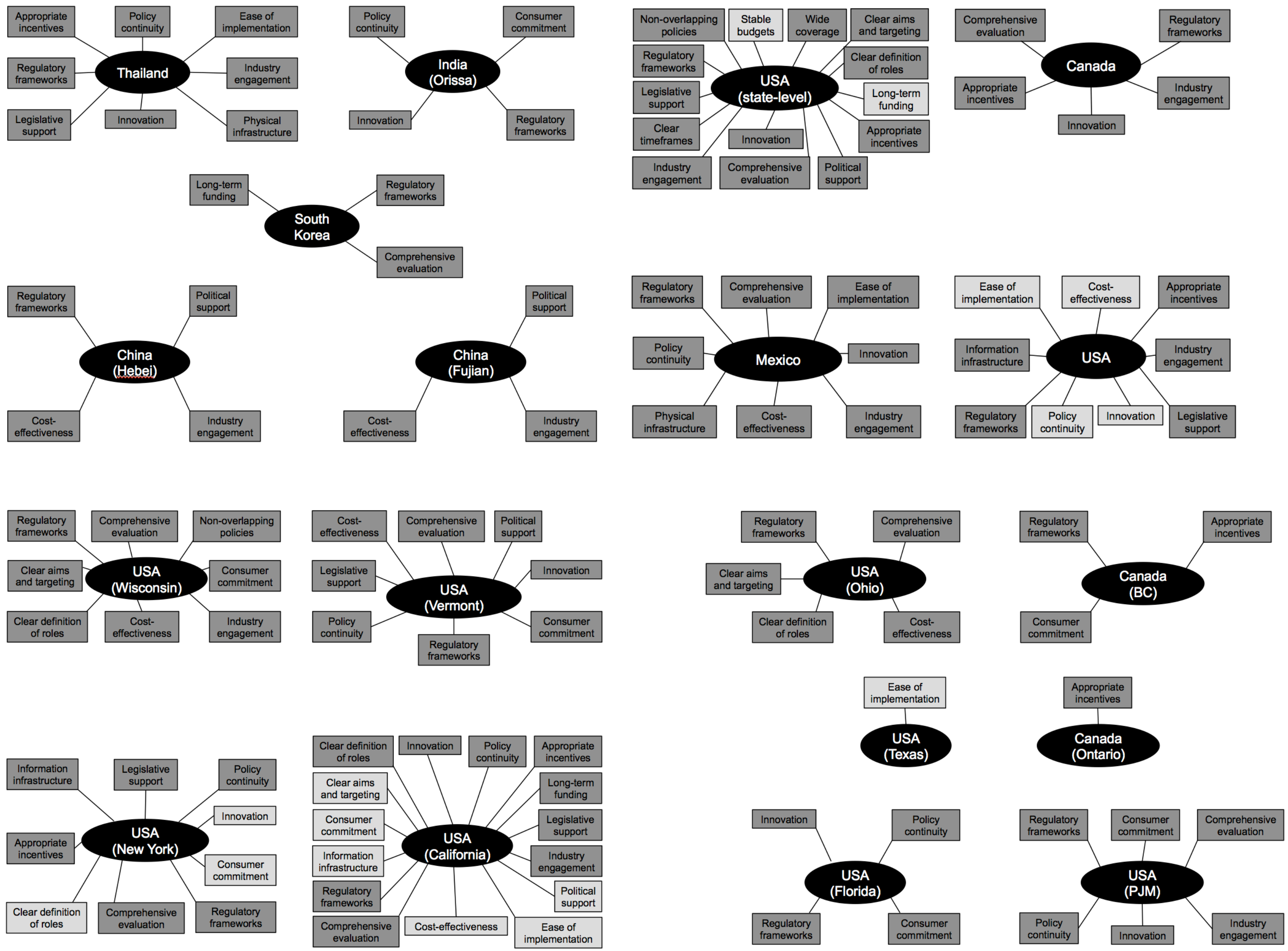
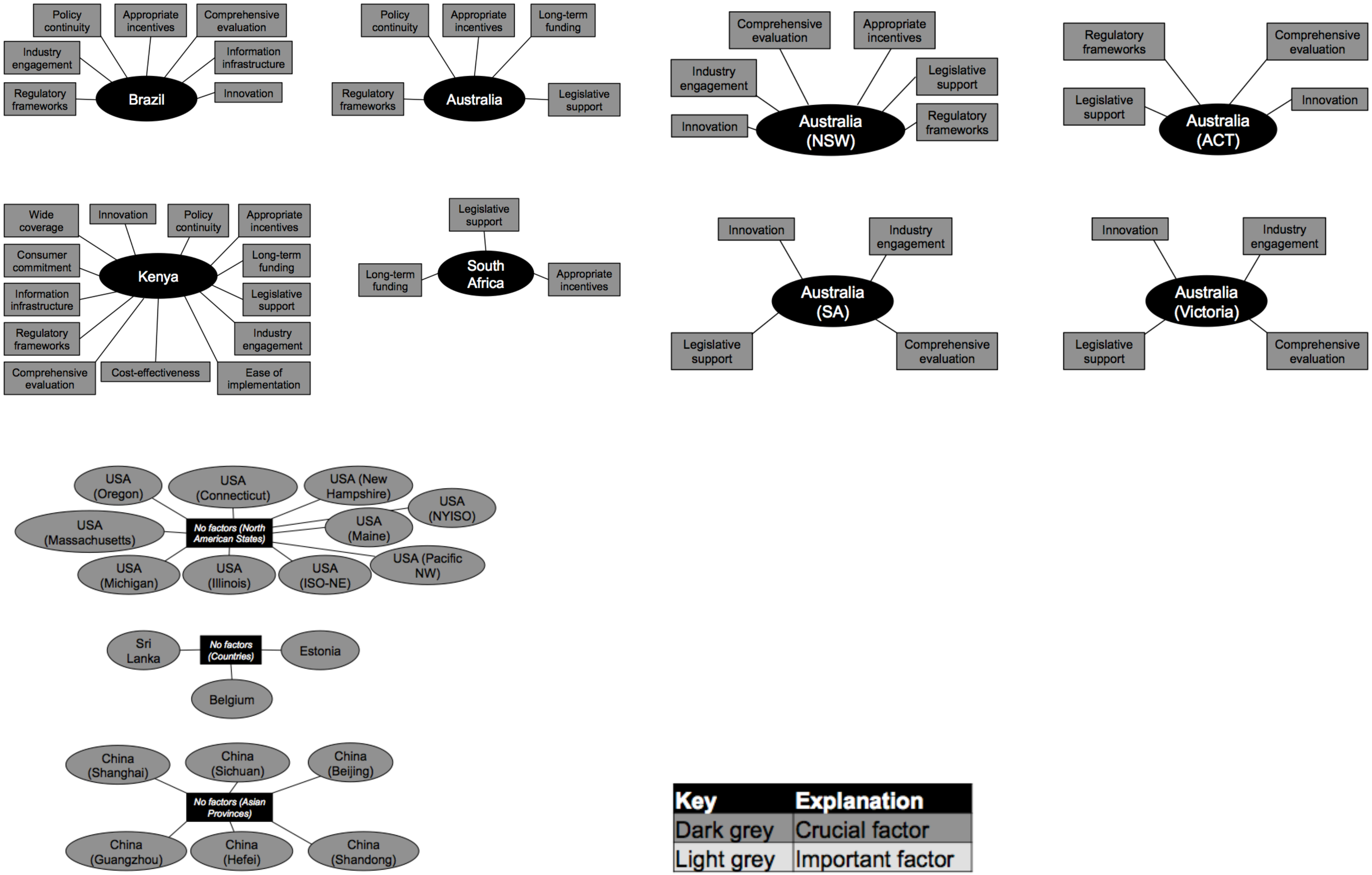


Figure 41 (continued)



Key	Explanation
Dark grey	Crucial factor
Light grey	Important factor

Figure 42: the key failure factors by country/state based on the combined analysis results

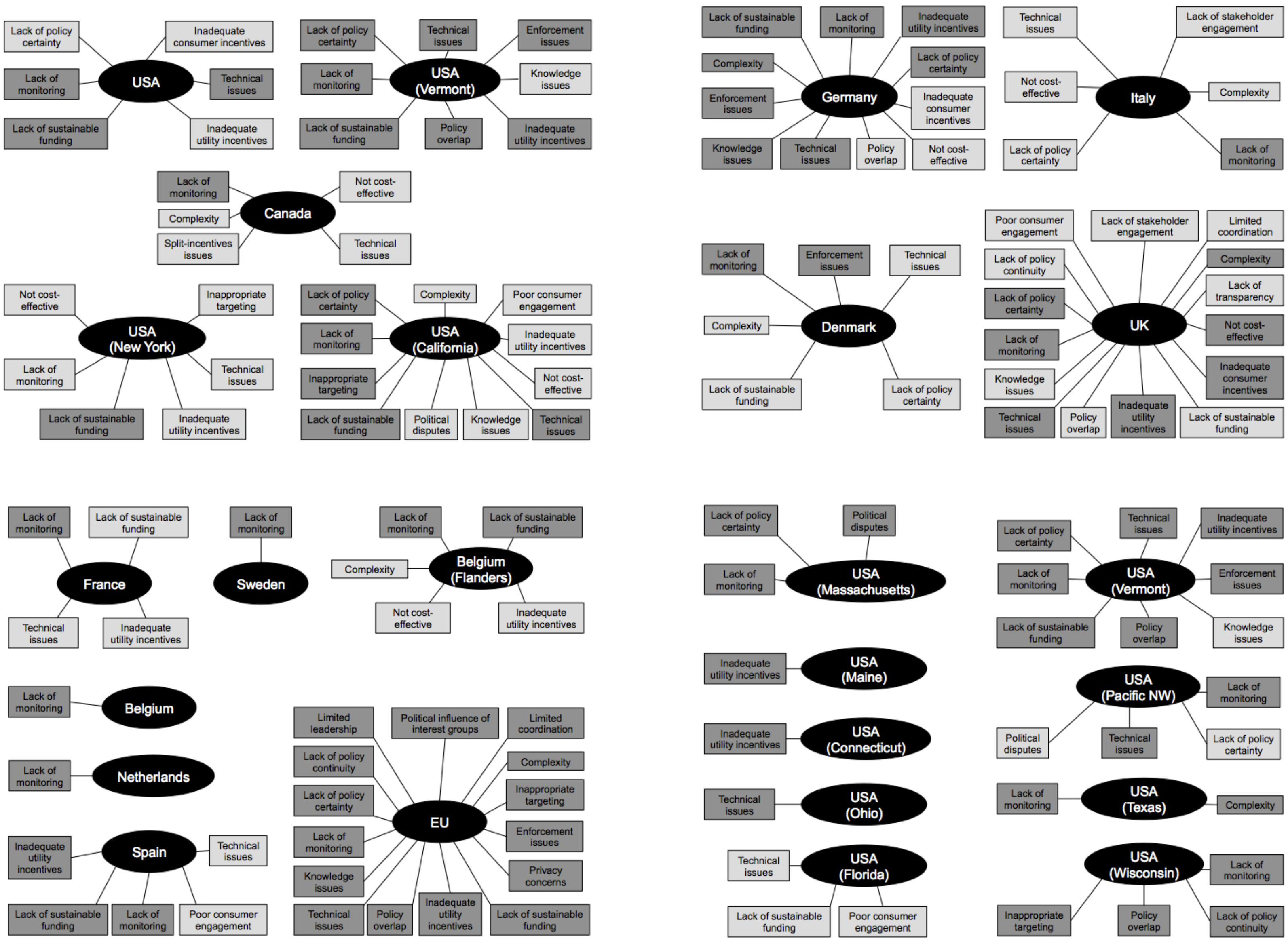


Figure 42 (continued)

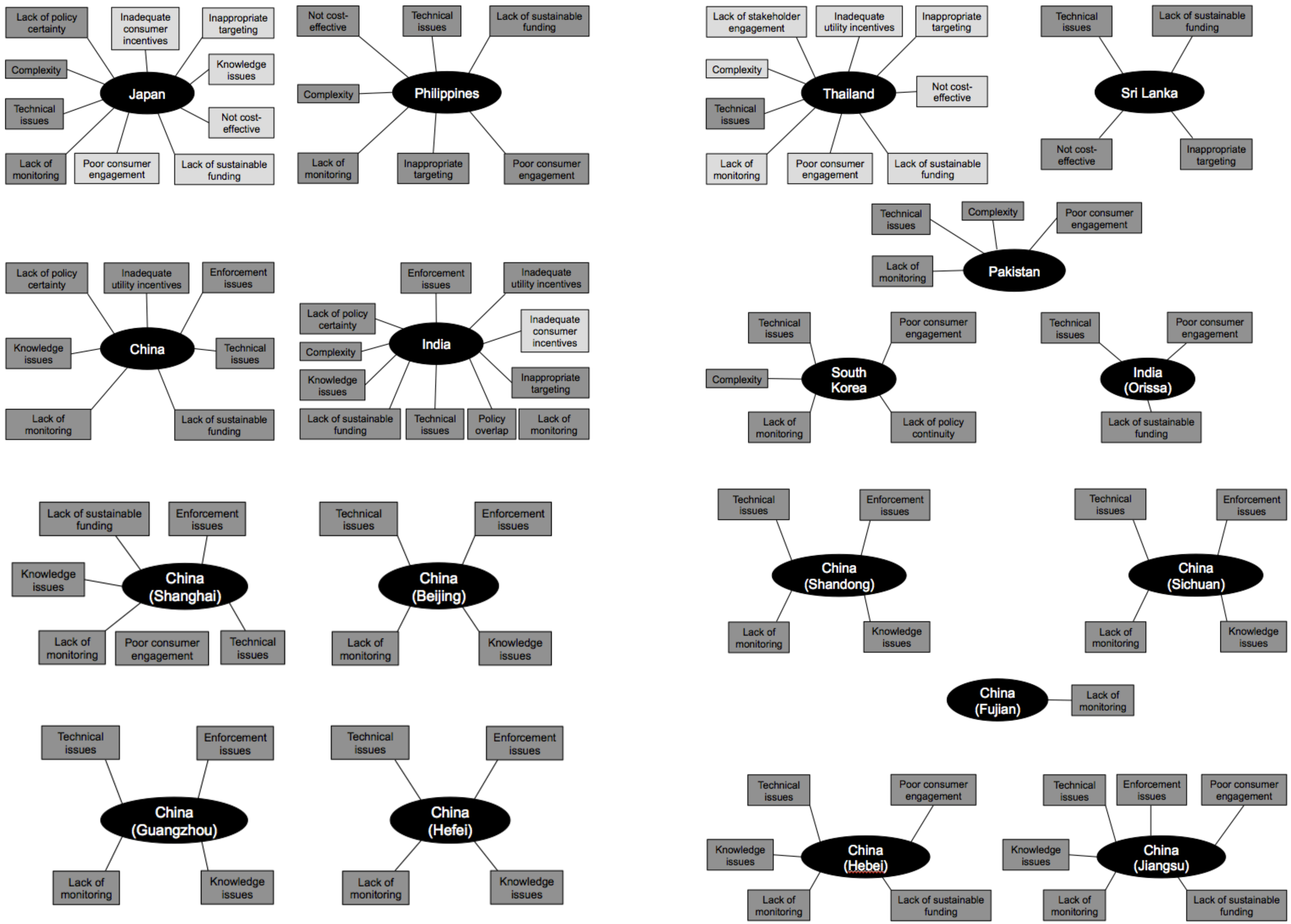
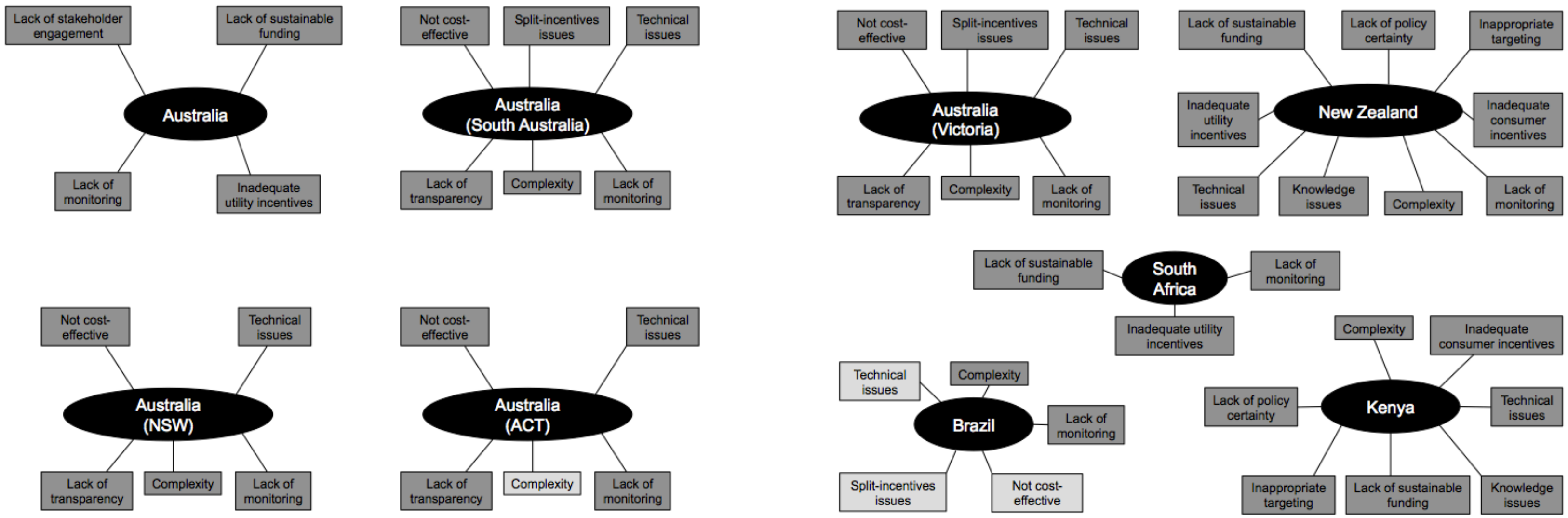
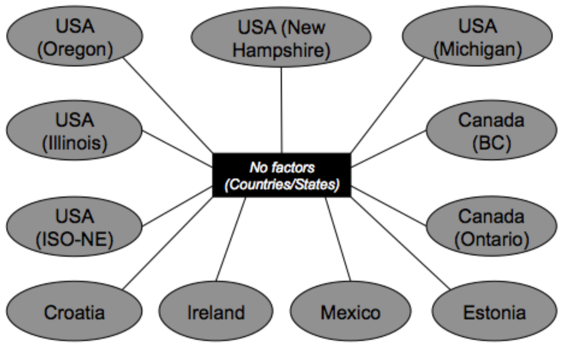


Figure 42 (continued)





Key	Explanation
Dark grey	Crucial factor
Light grey	Important factor

Table 15 summarises the top success and failure factors overall number countries/states where they are ‘crucial’ or ‘important’ fa combined analysis. The results are grouped by continent.

Continental Group	Success Factors	Number of Countries/ States	Failure Factors
Global	Regulatory frameworks	35	Lack of monitoring
	Innovation	28	Technical issues
	Comprehensive evaluation	27	Lack of sustainable funding
	Appropriate incentives	27	Complexity
	Industry engagement	26	/
North America	Regulatory frameworks	12	Lack of monitoring
	Comprehensive evaluation	9	Technical issues
	Innovation	9	Inadequate utility incentives
	Appropriate incentives	7	Lack of policy certainty
	Policy continuity	7	Lack of sustainable funding
	Industry engagement	7	/
	Consumer commitment	7	/
Europe	Regulatory frameworks	11	Lack of monitoring
	Appropriate incentives	11	Lack of sustainable funding
	Comprehensive evaluation	10	Technical issues
	Legislative support	8	Complexity
	Industry engagement	7	Inadequate utility incentives
	Innovation	7	/
Asia	Regulatory frameworks	7	Technical issues
	Industry engagement	7	Lack of monitoring
	Innovation	6	Knowledge issues
	Ease of implementation	5	Lack of sustainable funding
	Cost-effectiveness	5	Enforcement issues
	Policy continuity	5	Poor consumer engagement
AUS, AFR, SA	Legislative support	7	Lack of monitoring
	Comprehensive evaluation	6	Complexity
	Innovation	6	Technical issues
	Regulatory frameworks	5	Not cost-effective
	Appropriate incentives	5	Lack of transparency
	Industry engagement	5	/

Table 15: the key success and failure factors by cont

As table 15 shows, the top success factors across DSM policies are: regulatory frameworks, innovation, comprehensive evaluation, appropriate incentives and industry engagement. Of these, regulatory frameworks is the most important success factor, as it appeared as the most important factor in over half of the countries/states under examination (35/66 countries/states). The top failure factors across DSM policies are: a lack of monitoring, technical issues, a lack of sustainable funding and complexity. Of these, a lack of monitoring is the most important failure factor, as it appeared as the most important factor in two-thirds of the policies under examination (44/66 countries/states).

If the overall results by country/state (averaged across DSM policies) are compared with the overall results by policy (averaged across countries/states) (which were presented in sub-section 5.2.5), the key success factors that overlap between the analyses are regulatory frameworks and appropriate incentives, and the key failure factors that overlap are a lack of monitoring and technical issues (primarily programme management issues). Thus, the concise answer to the first part of research question two is that DSM policies succeed where regulatory frameworks and appropriate incentives are in place, and fail where there is a lack of monitoring and technical issues related to programme administration are present. In order to answer the second part of research question two, section 5.3 identifies which DSM policies have been successful (and unsuccessful) in particular countries/states.

5.3 Successful DSM Policies

The individual evaluations included in the systematic review determine the impacts and success of certain DSM policies in specific countries/states, so the purpose of this research is to look for high-level patterns through synthesis and aggregation in order to reach generality and representation for the success of various DSM policies. As previously justified in chapters 1-3, this is an understudied area. In order to determine the general success of each DSM policy across countries/states, qualitative statements on policy success are taken from all of the relevant evaluations in the systematic review sample for each policy within a particular country/state. As discussed and justified in sub-section 5.2.1, a scale of 1-5 is used, which was shown in figure 28 and is reproduced below:

1 = A failed policy that met none of its original objectives
2 = A poorly performing policy that met few of its original objectives
3 = An average performing policy that met most of its original objectives
4 = A policy that performed well and met all of its original objectives
5 = A highly successful policy that performed beyond its original objectives

The *Policy Success Weighting Scale* draws parallels to the *Factor Weighting Scale* (shown in figure 27), where certain words and phrases are converted into quantitative scores on the scale:

1 = ‘failed’, ‘unsuccessful’, ‘ill-fated’, ‘ineffective’, or equivalent
2 = ‘less successful’, ‘performed poorly’, ‘few successes’, ‘less effective’, or equivalent
3 = ‘average performance’, ‘satisfactory’, ‘met most of the objectives’, or equivalent
4 = ‘performed well’, ‘met all of the objectives’, ‘successful’, ‘effective’, or equivalent
5 = ‘highly successful’, ‘highly effective’, ‘performed beyond objectives’, or equivalent

These phrases were extracted directly from the evaluations in the production of the scale. Thus, the scale was determined inductively. Although there is arguably a bias in the judgement of the evaluators within the evaluations as to how the policy in question performed, the documents are of a high-quality and thus the expert judgement of the evaluators from conducting objective, high-quality evaluations should be considered reliable. Despite this, as discussed in sub-section 5.2.1, the evaluators’ use of words may vary. For example, how one evaluator uses the phrase ‘average performance’ may differ from another evaluator’s use of the same phrase. This is a challenge faced when examining the underlying mechanisms of interventions and policies, and is an area for further methodological research.

The first time the systematic review results were analysed, all of the documents were consulted and the success factors, failure factors and words or phrases related to policy success were noted. The factors and phrases were then put into groups within the scales. The systematic review results were then analysed a second time so that the relevant information from the evaluations could be categorised. The second iteration was needed to overcome the methodological problem of where phrases and factors identified in later documents could not be applied to earlier documents (as discussed in sub-section 5.2.1). Thus, by using the first iteration only to identify phrases and factors generally, rather than synthesising specific findings, more robust categorisations could be determined in the second iteration.

An average for each DSM policy was taken across all of the evaluations within each document for all countries/states. A figure to one decimal place was produced (but still within the 1-5 scale range above) for more detailed comparisons, as with the *Factor Weighting Scale*. This produced the Policy Success Weighting (PS_p) (where $_p$ is the policy in question) and the results are shown in table 16 below.

DSM Policy	**PS_p**
Utility Obligations (UO)	4.4
Performance Standards (PS)	4.2
Infrastructure Rollouts (IR)	4.0
Utility Business Models (UBM)	3.7
Research and Development (R&D)	3.5
Incentive Payment-Based Demand Response (IPBDR)	3.4
Loans and Subsidies (L&S)	3.4
Voluntary Programmes (VP)	3.3
Market Transformations (MT)	3.3
Price-Based Demand Response (PBDR)	3.3
Information Campaigns (IC)	3.1
Labelling (LB)	3.0
AVERAGE	**3.6**

Table 16: the overall success of different DSM policies

Policies that performed above the average PS_p of 3.6 are considered the most successful DSM policies. None of the policies score below 3.0, which suggests that none of them perform consistently poorly and instead show examples of both success and failure in different countries/states. Thus, success for these policies is context-specific. However, utility obligations, performance standards, infrastructure rollouts and alternative utility business models appear to perform consistently well across the countries/states where high-quality evaluations exist for the policies. Utility obligations and performance standards had large sample sizes across different countries/states. Although alternative utility business models had a large sample size, this was not geographically diverse and instead evaluations were primarily undertaken for state-level policies within the USA. The sample size for infrastructure rollouts was the smallest out of the twelve policies in the analysis and hence a larger sample size may influence its PS_p. The sample sizes for each policy were shown in table 8 and discussed in chapter four. At the other end of table 16, labelling, information campaigns, price-based demand response and market transformations are the least successful DSM policies overall.

The results show how the individual DSM policies compare with each other across contexts and countries/states. However, context is important and thus looking at the results by country/state gives a more comprehensive understanding of how successful the policies in table 16 have been. It is similarly important to note that different categories of DSM policy are usually implemented to meet different (though related) objectives, such as carbon emissions reduction, energy security and developing new markets, as summarised in figure 23 in sub-section 4.2.2. As a result, it is more fruitful to compare the overall performance of policies in different countries/states rather than comparing different categories of DSM policy generally (not in particular contexts). As Boza-Kiss *et al*. (2013) argue: “any of them can be cost-effective if selected, designed, implemented and enforced in a tailored way to local resource, capacities and cultures”.

To identify which policies have been successful in particular countries/states, all of the values for PS_p from each evaluation are listed by country/state and an average is taken for each policy in a given country/state. The method also allows the identification of unsuccessful policies for each country/state, which is equally interesting although it is not a necessary part of answering the second part of research question two.

The results convey that 41 countries/states (out of the 66 countries/states in the sample) have implemented at least one successful DSM policy (either as individual policies or policy packages). However, 47 countries/states have implemented at least one unsuccessful DSM policy (either as individual policies or policy packages). In the case of policy success, this refers to policies scoring 3.6-5.0 on the *Policy Success Weighting Scale*, whereas policies that score 1.0-3.5 represent those that perform below average and are considered unsuccessful.

The results are presented in two forms: firstly, the names of the countries/states that have experienced success with particular policies are given (figure 43), followed by a global map showing the number of successful policies by country/state (figure 44). The same visualisations are given for unsuccessful policies (figures 45 and 46). Policies are ranked by the total number of countries/states that have successfully (or unsuccessfully) implemented them.

UO	Belgium (Flanders), Italy, Japan, France, Brazil, Australia (New South Wales, Australian Capital Territory, South Australia, Victoria), Denmark, USA (state-level, Vermont), UK, USA, EU, Canada
PS	Denmark, USA (state-level, Vermont, California), China, UK, USA, EU, Australia
L&S	Thailand, USA (New York, California), Estonia, India (Orissa), Denmark, China, UK, USA
UBM	China (Hebei, Fujian), USA (New York, state-level, Vermont, California, Ohio), UK, USA
IPBDR	USA (New York, Florida, California), China, UK, USA, Spain
PBDR	USA (PJM region, Vermont, California), France, China, UK, USA
IC	Thailand, Denmark, Germany, USA (California), China, UK, South Korea
IC/L&S	USA (Illinois, Massachusetts, Wisconsin), Germany, China, USA
R&D	Denmark, USA (California), China, UK, USA
IPBDR/PBDR	USA (PJM region, NYISO region, ISO-NE region), China (Jiangsu, Beijing)
UBM/MT	USA (New York, Pacific Northwest region, Massachusetts, California), USA
MT	Thailand, USA (California), Sweden
IR	USA (California), UK, Australia
LB	Thailand, Denmark, China
VP	Denmark, China
PS/IC	USA (Pacific Northwest region)
PS/LB/IC	Philippines
PS/LB	China

Figure 43: successful DSM policies by country/state

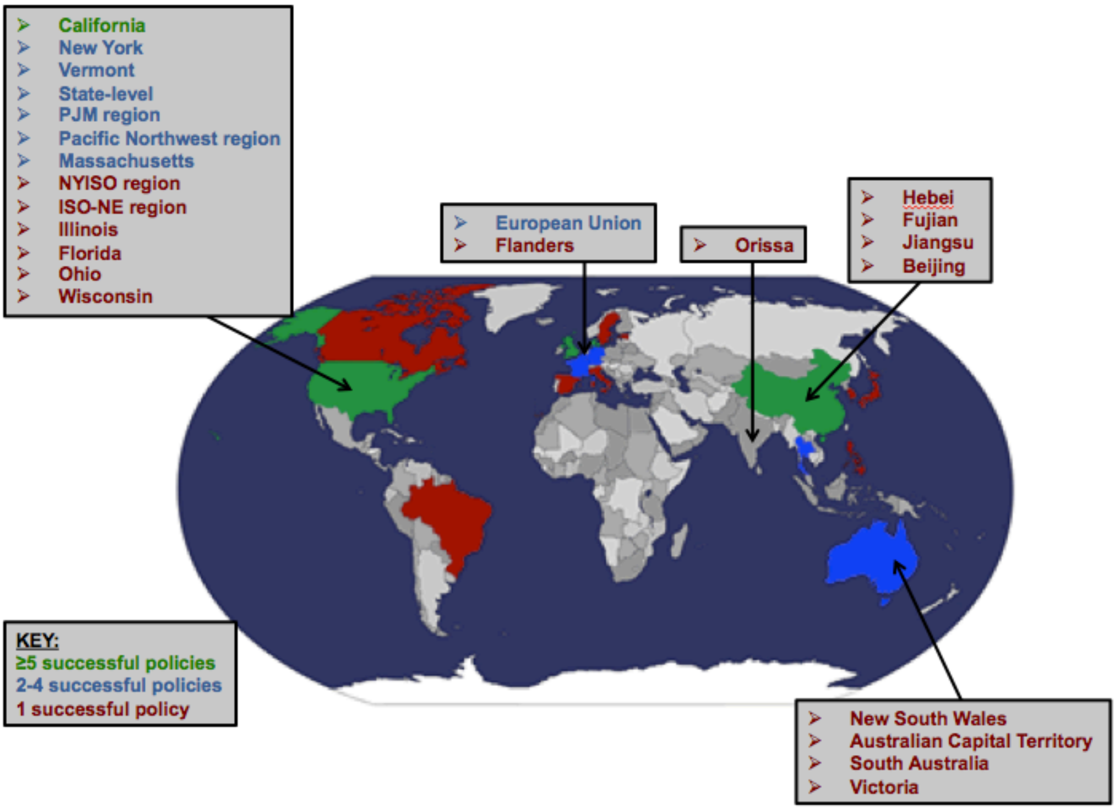


Figure 44: countries/states that have experienced DSM policy success

LB	Australia, Canada, Croatia, EU, India, Japan, Mexico, Netherlands, South Korea, Sweden, UK, USA
L&S	Canada, EU, India, Mexico, Netherlands, South Africa, South Korea, UK, USA (state-level, Oregon), Canada (Ontario)
IC	Australia, Croatia, EU, India, Indonesia, Ireland, Mexico, Netherlands, Pakistan, Sweden, USA
IPBDR	Australia, EU, India, Mexico, New Zealand, South Africa, South Korea, USA (Ohio), Canada (BC)
PS/LB	EU, Pakistan, China (Jiangsu, Shanghai, Beijing, Guangzhou, Hefei, Shandong, Sichuan)
PS	Canada, Croatia, Germany, India, Mexico, Netherlands, USA (New York)
IC/L&S	Mexico, New Zealand, Sri Lanka, USA (Illinois, Maine, Ohio, New Hampshire)
MT	Australia, Japan, Spain, UK, USA, USA (state-level)
UO	Australia, Netherlands, South Africa, USA (Oregon), Canada (Ontario)
UBM	China, Denmark, USA (Wisconsin, Michigan), Canada (Ontario)
PBDR	EU, South Africa, South Korea, USA (Texas), Canada (BC)
R&D	France, India, Mexico, Philippines, USA (Wisconsin)
IR	EU, Japan, USA, Canada (BC)
VP	Germany, South Korea, USA
IPBDR/PBDR	China, USA
PS/IC	Belgium, USA (Wisconsin)
PS/LB/UO/L&S	Italy, USA (California)
VP/L&S	India, UK
UBM/MT	USA (state-level)
IC/L&S/MT	USA (California)

Figure 45: unsuccessful DSM policies by country/state

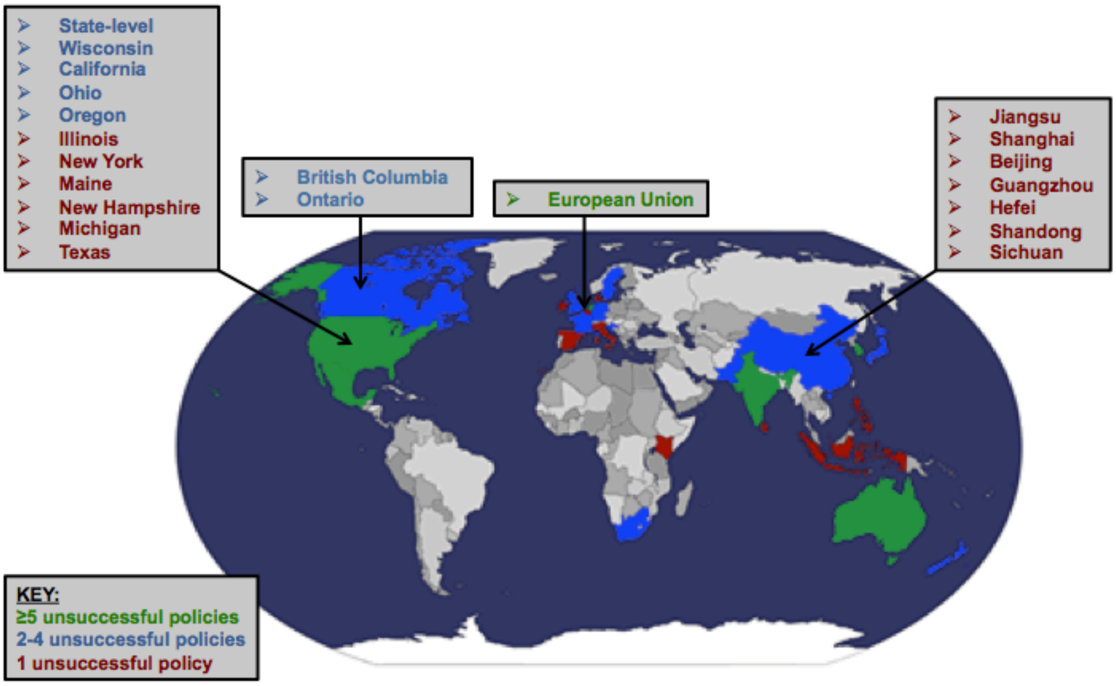


Figure 46: countries/states that have experienced DSM policy failure

The two global maps (figures 44 and 46) are colour-coded to show the number of successful or unsuccessful DSM policies that have been implemented with high-quality evaluations. Countries coloured in green have implemented ≥5 successful or unsuccessful policies (the same threshold as the *Factor Frequency Threshold* shown in figure 26), those coloured in blue have implemented 2-4 successful or unsuccessful policies and those coloured in red have implemented one successful or unsuccessful policy. The boxes show the states and provinces that have implemented successful or unsuccessful policies. The same colour coding system is used. A table with the DSM policies that have been successful or unsuccessful in these countries/states is included in the Appendix (Appendix Figures 6 and 7).

In terms of policy success, the results show that California and China have the greatest number of successful DSM policies, each having successfully implemented 10/21 policies (twelve individual policies and nine policy packages) included in the research. They are closely followed by the USA and the UK (each with 9/21 successful policies), and Denmark also performs well with 7/21 successful policies. Almost two thirds of the countries/states included in the systematic review (41/66) have implemented at least one successful DSM policy.

In terms of less successful policies, the results portray that the European Union (EU), India and Mexico have the greatest number of unsuccessful DSM policies, each with 7/21 unsuccessful policies. They are closely followed by the USA (with 6/21 unsuccessful policies) and Australia, the Netherlands and South Korea (each with 5/21 unsuccessful policies). With the exception of Mexico and the Netherlands, these countries have all implemented and evaluated at least one successful DSM policy as well. In some cases, such as the USA, the number of unsuccessful policies (6/21) is overshadowed by the greater number of successful policies (9/21). The USA has had less success at a national level with information-based or voluntary DSM policies than regulatory or market-based policies. Although some market-based policies, such as market transformations, appear not to have performed well, they have worked well when combined with alternative utility business models (such as system benefits charges to provide a long-term funding stream). Furthermore, some information-based policies, such as information campaigns, have also worked well in the USA as part of policy packages, particularly alongside loans and subsidies, rather than as stand-alone policies.

For the EU, the evidence base shows that it has been less successful in stimulating demand response policies (IPBDR and PBDR), infrastructure rollouts (though this could change in the future following the implementation of the *Smart Meter Rollout Directive*), labelling, loans and subsidies, information campaigns and the policy package of performance standards with labelling. However, the EU has been successful at stimulating utility obligations and performance standards. In contrast, the two other countries/states with the greatest number of unsuccessful DSM policies, India and Mexico, have not experienced success with any of the individual policies or policy packages. The Indian state of Orissa (now called Odisha) has experienced some success with loans and subsidies, but national level policies have generally been unsuccessful. Despite this, it is important to note that the evidence base of high-quality evaluations for India is limited, so conclusions are drawn from a smaller sample. Despite this, the literature review and the quick scoping review conducted prior to the systematic review identified that India has had limited experience with implementing demand-side policies, instead focussing on supply-side policies.

Two further short case studies that are interesting to examine are the US states of California and New York. Both have strong evidence bases in terms of the number of high-quality documents and evaluations (as shown in table 11 in chapter four) and the number of DSM policies implemented. Furthermore, they have a high number of successfully implemented policies (10/21 and 4/21 policies respectively). However, unlike the USA at a national level and the UK, the number of unsuccessful policies for these states is low. The USA and the UK both have high numbers of unsuccessful policies (6/21 and 4/21 policies respectively) alongside a high number of successful policies. For California, a clear pattern emerges. The evidence shows that it has been more successful with individual DSM policies than policy packages, as nine out of its ten successful policies are individual policies and the two unsuccessful policies are more complex policy packages involving more than two policies (IC/L&S/MT

and PS/LB/UO/L&S). Performance standards is the only policy that has been unsuccessful in the sample for New York (as discussed in sub-section 5.2.6).

Overall, in comparing the results for successful and unsuccessful policies, it appears that the USA, China and west-Europe have the largest number of successful policies, whereas Asia and Europe more broadly have a large number of unsuccessful policies. Other regions, such as Australasia, South America and Africa appear to fall between the two groups and show a few examples of successful DSM policies but also a few examples of unsuccessful policies. For example, four Australian states have had success with utility obligations (New South Wales, Australian Capital Territory, South Australia and Victoria) and Australia at a national level has been successful with policies for infrastructure rollouts and performance standards. However, Australia (at a national level) has been unsuccessful with policies for incentive payment-based demand response, market transformations, utility obligations, labelling and information campaigns. Nevertheless, the evidence base in Australasia, South America and Africa is smaller than that of North America, Europe and Asia, and thus conclusions are drawn from a smaller sample.

In comparing the results in figures 43-46 to those summarised in table 16, five of the top seven policies by successful implementation overlap between analyses (utility obligations, performance standards, alternative utility business models, incentive payment-based demand response and loans and subsidies). The top seven databases are examined (rather than the top five or the top ten databases) for consistency of discussion with the top seven databases in sub-section 3.4.7. Similarly, four of the top seven policies by unsuccessful implementation overlap between analyses (labelling, information campaigns, loans and subsidies and incentive payment-based demand response). Loans and subsidies and incentive payment-based demand response feature in the top seven policies for both successful and unsuccessful implementation, which is reflected in table 16 by the policies sitting in the middle of the table with policy success scores of 3.4. As previously argued, it is more useful to compare the success of different DSM policies by the number of countries/states that have successfully or unsuccessfully implemented them, in order to overcome issues of comparing the performance of policies that were implemented to meet different policy objectives.

5.4 Research Question 2 Conclusions

This chapter aimed to answer research question two: how and why do demand-side management policies succeed or fail, and what policies have been successful? Although it is a complex question to unravel and forms the central part of the thesis, the overall conclusion is that across countries/states and DSM policies the success and failure factors shown below are the most crucial:

Overall success factors:

- Regulatory frameworks
- Appropriate incentives

Overall failure factors:

- A lack of monitoring
- Technical issues (programme management issues)

The above two success factors and two failure factors are the results when averaged across both analyses (the combined frequency-weighting analysis by DSM policy when the results are averaged across countries/states and the combined frequency-weighting analysis by country/state when the results are averaged across DSM policies).

The results by DSM policy are listed below (i.e. the number of DSM policies that have the following specific factors as ‘crucial’ or ‘important’ factors):

Success factors by DSM policy:

- Regulatory frameworks (6/12 policies)
- Legislative support (4/12 policies)
- Appropriate incentives (3/12 policies)
- Information infrastructure (3/12 policies)
- Consumer commitment (2/12 policies)

Failure factors by DSM policy:

- Technical issues (programme administration issues) (6/12 policies)
- A lack of policy certainty (3/12 policies)
- A lack of monitoring (2/12 policies)
- Inadequate utility incentives (2/12 policies)
- Inadequate consumer incentives (2/12 policies)

The results by country/state are listed below (i.e. the number of countries/states that have the following specific factors as ‘crucial’ or ‘important’ factors):

Success factors by country/state:

- Regulatory frameworks (35/66 countries/states)
- Innovation (28/66 countries/states)
- Comprehensive evaluation (27/66 countries/states)
- Appropriate incentives (27/66 countries/states)
- Industry engagement (26/66 countries/states)

Failure factors by country/state:

- A lack of monitoring (44/66 countries/states)
- Technical issues (programme administration issues) (39/66 countries/states)
- A lack of sustainable funding (25/66 countries/states)
- Complexity (22/66 countries/states)

An extension to the main analysis explored whether or not there are significant associations between success factors and between failure factors. *Pearson’s product-moment correlation coefficient* (*r*) was used and the following success and failure factors show the most associations and the least associations with other factors:

Success factors with the most associations:

- ➢ Legislative support (LS) (14 factor associations)
- ➢ Regulatory frameworks (RF) (13 factor associations)
- ➢ Non-overlapping policies (NO) (11 factor associations)

- Appropriate incentives (AI) (10 factor associations)
- Clear aims and targeting (CA) (10 factor associations)

Success factors with the least associations:

- Long-term funding (LT) (0 factor associations – 0.0% of the sample)
- Stable budgets (SB) (0 factor associations – 0.0% of the sample)
- Return on investments (RI) (0 factor associations – 0.0% of the sample)
- Innovation (IV) (0 factor associations – 0.0% of the sample)

Failure factors with the most associations:

- Policy overlap (PO) (15 factor associations – 60.0% of the sample)
- Lack of transparency (LT) (14 factor associations – 56.0% of the sample)
- Complexity (CX) (13 factor associations – 52.0% of the sample)
- Lack of monitoring (LM) (13 factor associations – 52.0% of the sample)

Failure factors with the least associations:

- Lack of policy certainty (LC) (0 factor associations – 0.0% of the sample)
- Utility opposition (UO) (0 factor associations – 0.0% of the sample)
- Technical issues (TI) (0 factor associations – 0.0% of the sample)

The *Success Factor Association Scale* was developed to show the proportion (%) of the other factors that a given factor is significantly associated with (at the 0.05 significance level). The scale identified the factors that are important only in the context of specific other factors being present.

Globally, the evidence base shows that the following DSM policies have been the most successful and the least successful in the past overall:

Most successful DSM policies:

- Utility obligations
- Performance standards (for equipment, appliances, and buildings)
- Infrastructure rollouts
- Alternative utility business models

Least successful DSM policies:

- Labelling
- Information campaigns
- Price-based demand response
- Market transformations

The evidence base is clear in showing that utility obligations, performance standards, infrastructure rollouts and alternative utility business models have been the most successful DSM policies, and labelling, information campaigns, price-based demand response and market transformations have been the least successful DSM policies. If the same method is applied to analyse the number of countries/states that have been successful or unsuccessful with various DSM policies, the following policies overlap with the results from the overall analysis:

Most successful DSM policies:

- Utility obligations
- Performance standards
- Alternative utility business models
- Incentive payment-based demand response
- Loans and subsidies

Least successful DSM policies:

- Labelling
- Information campaigns
- Loans and subsidies
- Incentive payment-based demand response

Here, loans and subsidies and incentive payment-based demand response feature as both successful and unsuccessful policies, which is reflected in their Policy Success Weightings (PS_p) of 3.4 (average policy success scores). This highlights that, compared with utility obligations, performance standards and alternative utility business models, which appear to be more universally successful, loans and subsidies and incentive payment-based demand response are more context-specific and show examples of success in some countries/states and failure in other countries/states. The main argument in this respect is that it is more useful to compare the success of different DSM policies by the number of countries/states that have successfully or unsuccessfully implemented them, as an overall (country/state-independent) comparison does not acknowledge that different types of DSM policy are often implemented to meet different policy objectives.

This generally matches the findings of the few studies that have been conducted in this area, such as Ürge-Vorsatz *et al*. (2007) and Harmelink *et al*. (2008). For example, Ürge-Vorsatz *et al*. (2007) found that performance standards (particularly appliance standards and building codes), loans and subsidies (particularly tax exemptions or reductions), utility obligations and labelling perform the best in terms of cost-effectiveness and carbon emissions reductions. Although this thesis found that loans and subsidies had an average policy success score, it was the most diverse in terms of performance across countries/states (with incidences of both success and failure). Furthermore, although the specific DSM policy level was not analysed (e.g. 'tax exemptions' within the broader category of 'loans and subsidies', as per table 2 in chapter two), it was found that tax incentives generally performed better than subsidy policies in the sample. However, in contrast to Ürge-Vorsatz *et al*. (2007), the findings show that labelling policies have generally not performed well overall.

In addition to identifying the most and least successful DSM policies, it is similarly interesting to look at the most and least successful countries/states in implementing DSM policies, as shown below:

Most successful countries/states:

- ❖ California (USA) (10/21 policies successfully implemented)
- ❖ China (10/21 policies successfully implemented)
- ❖ UK (9/21 policies successfully implemented)
- ❖ USA (9/21 policies successfully implemented)

Least successful countries/states:

- ❖ European Union (EU) (7/21 policies unsuccessfully implemented)
- ❖ India (7/21 policies unsuccessfully implemented)
- ❖ Mexico (7/21 policies unsuccessfully implemented)

- USA (6/21 policies unsuccessfully implemented)

California, China, the USA and the UK have experienced success with the full range of DSM policy categories from demand response policies and research and development programmes to performance standards and alternative utility business models. In contrast, the evidence base shows that the EU, India, Mexico and the USA have experienced policy failure with a large range of DSM policy categories. The breakdown of successful and unsuccessful DSM policies by country/state is summarised in the Appendix (Appendix Figures 6 and 7). In the case of the USA, its experience with DSM policy is the greatest of any country/state, which explains why it has both a high number of successful and unsuccessful policies.

6 Chapter 6: Conclusion

6.1 Research Findings

6.1.1 Research Overview

This thesis aimed to determine the mechanisms behind the success and failure of demand-side management (DSM) policies. DSM refers to activities on the demand-side of energy meters that seek to meet energy policy objectives, such as energy security and carbon emissions reduction. There has been much research on DSM since the energy crises of the 1970s, but limited attention has been given to the policy side of DSM, particularly in the academic literature. DSM policy evaluations that have been conducted tend to focus on policy impacts rather than the key factors that determine policy success or failure.

This thesis aimed to contribute to filling these research gaps by answering the following research questions:

1. What DSM policies have been implemented around the world with high quality documented evaluations?
2. How and why do DSM policies succeed or fail, and what policies have been successful?

To answer the research questions, a global systematic review of the DSM policy evaluation evidence base was undertaken. Only high-quality evaluations were included in the systematic review, which were determined through an assessment of study quality, and the final sample of 119 documents included 102 academic, industrial and policy papers and reports, and 17 interviews with DSM experts using multi-criteria decision-making (MCDM) analysis. The mixed methods realist synthesis type of systematic review was utilised, as it focuses on programme mechanisms and MCDM analysis was undertaken as a secondary method to form one of the systematic review databases and to act as a form of cross-validation for the systematic review. The systematic review sample covered 35 databases: nine academic databases (25 documents), eleven industrial databases (59 documents), thirteen policy databases (14 documents), one MCDM analysis expert database (17 interviews) and one review update database (4 documents). The key findings for the two research questions are given in the following sub-sections.

6.1.2 Key Findings: Research Question 1

30 countries and 36 states (including regions and provinces) across six continents were included in the research and were determined inductively from the sample. The majority of documents focused on the USA (25 documents), the UK (22 documents), California (USA) (20 documents), France (13 documents) and China (12 documents). Twelve DSM policy categories and nine policy packages were included in the research:

DSM Policy Categories:

- IPBDR (Incentive payment-based demand response)
- PBDR (Price-based demand response)
- MT (Market transformations)
- IR (Infrastructure rollouts)
- UO (Utility obligations)
- LB (Labelling)
- PS (Performance standards)
- L&S (Loans and subsidies)
- UBM (Utility business models)
- R&D (Research and development programmes)
- IC (Information campaigns)
- VP (Voluntary programmes)

DSM Policy Packages:

- IPBDR/PBDR (policy package of *Incentive payment-based demand response* and *Price-based demand response*)
- UBM/MT (policy package of *Utility business models* and *Market transformations*)

- IC/L&S (policy package of *Information campaigns* and *Loans and subsidies*)
- PS/IC (policy package of *Performance standards* and *Information campaigns*)
- PS/LB/IC (policy package of *Performance standards*, *Labelling* and *Information campaigns*)
- PS/LB (policy package of *Performance standards* and *Labelling*)
- IC/L&S/MT (policy package of *Information campaigns*, *Loans and subsidies* and *Market transformations*)
- PS/LB/UO/L&S (policy package of *Performance standards*, *Labelling*, *Utility obligations* and *Loans and subsidies*)
- VP/L&S (policy package of *Voluntary programmes* and *Loans and subsidies*)

The majority of the documents in the sample included evaluations of more than one policy or country/state. Thus, the total number of policy evaluations within the 102 written documents (excluding the 17 MCDM analysis interviews) was 690. The evidence base is dominated by evaluations of individual policies (645 evaluations) rather than policy packages (45 evaluations). The DSM policies with the greatest number of evaluations are:

- ➢ Utility business models (primarily performance targets, system benefits charges and decoupling policies) (122 evaluations)
- ➢ Information campaigns (118 evaluations)
- ➢ Loans and subsidies (100 evaluations)
- ➢ Utility obligations (89 evaluations)
- ➢ Performance standards (for appliances, equipment and buildings) (81 evaluations)

At the other end of the scale, voluntary programmes (12 evaluations) and infrastructure rollouts (4 evaluations) had the least number of policy evaluations. However, the evidence base for infrastructure rollouts is likely to increase post-2020 following the implementation of the European Union (EU)'s *Smart Meter Rollout Directive* (*Directive 2009/72/EC*), which mandates that all countries in the EU must reach at least an 80% rollout of smart meters to small consumers by 2020.

The dominant policy package is information campaigns in combination with loans and subsidies (13 evaluations), followed by performance standards in combination with labelling (10 evaluations). In contrast, mixing these two combinations of policies together is the least frequent policy package (performance standards in combination with labelling and information campaigns) with just one evaluation (in the Philippines). The small number of evaluations focusing on policy packages could be explained by countries/states choosing to implement DSM policies individually, which usually entail lower administrative costs to governments, or by evaluation challenges. In the case of the latter, it is more complex to separate out the impacts for different parts of a policy package and there is less experience in the DSM policy evaluation literature on the practicalities of evaluating policy packages (as discussed in section 7.4). It is likely that both resource constraints and evaluation challenges contribute to explaining the finding.

The USA, California, the UK and China have the greatest diversity of implementation of DSM policies. The USA has implemented and evaluated fifteen different DSM policies and policy packages and California, the UK and China have each implemented and evaluated twelve different DSM policies and policy packages. The EU and Denmark also perform well in this respect having implemented and evaluated nine and eight different DSM policies and policy packages respectively.

The analysis also examined the primary reasons for governments to implement DSM policies. Twenty reasons were identified inductively from the sample and the top objectives (in terms of the number of evaluations that state them) were: to reduce carbon emissions (47 evaluations), to ensure energy security (41 evaluations), to increase the uptake of energy efficient technologies (36 evaluations) and to create new resources and markets (33 evaluations).

Although it is not an essential part of answering research question one, the evaluations were categorised by decade to look at temporal patterns. The results show that the number of high-quality DSM policy evaluations increases over time, particularly in recent decades (2000s and 2010s). This is especially the case in North America, Europe and east-Asia. However, limited digitisation of evaluations conducted in the 1970s-1990s could explain the findings and section 6.4 discusses how further research could explore this.

6.1.3 Key Findings: Research Question 2

Answering research question two forms the central part of the thesis. Overall, across DSM policies and countries/states, the following success and failure factors are crucial:

Overall success factors:

- Regulatory frameworks
- Appropriate incentives

Overall failure factors:

- A lack of monitoring (for evaluation)
- Technical issues (primarily programme administration issues)

The success and failure factors were determined using frequency analysis, weighting analysis and a combined frequency-weighting analysis. The above factors are the results from the combined analysis, which identified factors that were both frequent in the sample and highly weighted by evaluators. The proposed method for examining policy mechanisms utilises a number of thresholds, scales and equations, and the three main steps for the analysis are listed below. The specific details on the scales and equations were given in chapter five.

1) *Step 1:* determine the frequency of the factor in question by summing the number of evaluations that mention it as a factor (if ≥5 evaluations, it is considered frequent)
2) *Step 2:* determine the weighting of the factor in question by weighting the factor in each evaluation that discusses it based on each evaluator's

judgement and then take an overall average for the sample (if the overall weighting is 2.5-3.0, it is considered highly weighted)

3) *Step 3:* combine the two results using the *Combined Frequency-Weighting Equation* (if ≥5.0% of the theoretical maximum score, it is considered a key factor)

The same analytical process was used to identify the key success and failure factors by DSM policy (averaged across countries/states) and the key factors by country/state (averaged across DSM policies). The key success and failure factors by DSM policy are summarised in table 17 overleaf. Factors shown in bold are 'crucial' factors and those not shown in bold are 'important' factors.

Although the key success and failure factors were also examined for each of the 66 countries/states included in the systematic review, interesting observations at a continental level can be made. The key success and failure factors by continent are shown in table 18.

DSM Policy	Key Success Factors	Key Failure Factors
Incentive payment-based demand response	Regulatory frameworks	Technical issues
		A lack of policy certainty
Price-based demand response	**Regulatory frameworks**	Technical issues
	Appropriate incentives	A lack of monitoring
	Information infrastructure	Inadequate utility incentives
	Clear aims and targeting	Inadequate consumer incentives
	Consumer commitment	
Market transformations	Appropriate incentives	Technical issues
		A lack of monitoring
		A lack of policy continuity
Infrastructure rollouts	**Regulatory frameworks**	**Technical issues**
	Legislative support	**Limited coordination**
	Appropriate incentives	**A lack of transparency**
	Information infrastructure	**Political disputes**
	Industry engagement	**A lack of policy certainty**
	Consumer commitment	**Inadequate utility incentives**
	Clear aims and targeting	**Limited leadership**
	Political support	**Policy overlap**
	Physical infrastructure	Privacy concerns
	Non-overlapping policies	Negative public perception
	Policy continuity	Knowledge issues
		A lack of policy continuity
Utility obligations	**Regulatory frameworks**	*No factors discussed*
	Legislative support	
	Comprehensive evaluation	
	Clear definition of roles	
	Cost-effectiveness	
Labelling	Information infrastructure	Technical issues
Performance standards	Regulatory frameworks	*No factors discussed*
	Legislative support	
Loans and subsidies	*No factors discussed*	*No factors discussed*
Utility business models	*No factors discussed*	*No factors discussed*
Research & development programmes	Regulatory frameworks	**Technical issues**
	Legislative support	Inadequate consumer incentives
Information campaigns	*No factors discussed*	*No factors discussed*
Voluntary programmes	**Clear timeframes**	*No factors discussed*

Table 17: the key success and failure factors by DSM policy

Continent	Key Success Factors	Key Failure Factors
North America	Regulatory frameworks	Lack of monitoring
	Comprehensive evaluation	Technical issues
	Innovation	Inadequate utility incentives
	Appropriate incentives	Lack of policy certainty
	Policy continuity	Lack of sustainable funding
	Industry engagement	/
	Consumer commitment	/
Europe	Regulatory frameworks	Lack of monitoring
	Appropriate incentives	Lack of sustainable funding
	Comprehensive evaluation	Technical issues
	Legislative support	Complexity
	Industry engagement	Inadequate utility incentives
	Innovation	/
Asia	Regulatory frameworks	Technical issues
	Industry engagement	Lack of monitoring
	Innovation	Knowledge issues
	Ease of implementation	Lack of sustainable funding
	Cost-effectiveness	Enforcement issues
	Policy continuity	Poor consumer engagement
Australasia, Africa, and South America	Legislative support	Lack of monitoring
	Comprehensive evaluation	Complexity
	Innovation	Technical issues
	Regulatory frameworks	Not cost-effective
	Appropriate incentives	Lack of transparency
	Industry engagement	/

Table 18: the key success and failure factors by continent

The ratio of ‘crucial’ to ‘important’ factors is available at the country/state level rather than the continent level (thus it is not shown in table 18). The results from all three of the analyses (overall factors, factors by policy and factors by country/state) highlight the importance of regulatory frameworks and appropriate incentives as the most important success factors, and a lack of monitoring (for evaluation) and technical issues (primarily programmes management issues) as the most important failure factors.

An extension to the main analysis explored whether or not there were significant associations between success factors and between failure factors. *Pearson’s product-moment correlation coefficient* (*r*) was used, as the test examines statistical associations between variables. The purpose of the analysis was to test the significance of the associations between different success or failure factors in order to identify the clusters of factors that need to be present for DSM policies to be successful overall. The success and failure factors that had the most and least associations with other factors are shown below:

Success factors with the most associations:

- Legislative support (LS) (14 factor associations)
- Regulatory frameworks (RF) (13 factor associations)
- Non-overlapping policies (NO) (11 factor associations)
- Appropriate incentives (AI) (10 factor associations)
- Clear aims and targeting (CA) (10 factor associations)

Success factors with the least associations:

- Long-term funding (LT) (0 factor associations – 0.0% of the sample)
- Stable budgets (SB) (0 factor associations – 0.0% of the sample)
- Return on investments (RI) (0 factor associations – 0.0% of the sample)
- Innovation (IV) (0 factor associations – 0.0% of the sample)

Failure factors with the most associations:

- Policy overlap (PO) (15 factor associations – 60.0% of the sample)
- Lack of transparency (LT) (14 factor associations – 56.0% of the sample)
- Complexity (CX) (13 factor associations – 52.0% of the sample)
- Lack of monitoring (LM) (13 factor associations – 52.0% of the sample)

Failure factors with the least associations:

- Lack of policy certainty (LC) (0 factor associations – 0.0% of the sample)
- Utility opposition (UO) (0 factor associations – 0.0% of the sample)
- Technical issues (TI) (0 factor associations – 0.0% of the sample)

The *Success Factor Association Scale* was developed to show the percentage of the sample that a given factor is significantly associated with (at the 0.05 significance level). As discussed in chapter five, the scale did not rank factors individually by importance – this was the purpose of the combined frequency-weighting analysis. Instead, the scale identified the factors that are important only in the context of specific other factors being present.

The final aspect of answering research question two was the identification of successful DSM policies by country/state. Overall, the DSM policies shown below have been the most successful and the least successful across countries/states. A technique for determining DSM policy success was established using averaged policy weightings across the sample, as judged by the evaluators of each policy evaluation. The specific details of the techniques used were given in chapter five (section 5.3).

Most successful DSM policies:

- Utility obligations
- Performance standards
- Alternative utility business models
- Incentive payment-based demand response
- Loans and subsidies

Least successful DSM policies:

- Labelling
- Information campaigns
- Loans and subsidies
- Incentive payment-based demand response

Despite these findings, different DSM policies are usually implemented to meet different policy objectives and different countries/states have experienced success and failure with various DSM policies. However, the following countries/states have been the most successful and the least successful in implementing and evaluating a variety of different DSM policies:

Most successful countries/states:

- California (USA) (10/21 different DSM policies successfully implemented)
- China (10/21 different DSM policies successfully implemented)
- UK (9/21 different DSM policies successfully implemented)
- USA (9/21 different DSM policies successfully implemented)

Least successful countries/states:

- EU (7/21 different DSM policies unsuccessfully implemented)
- India (7/21 different DSM policies unsuccessfully implemented)
- Mexico (7/21 different DSM policies unsuccessfully implemented)
- USA (6/21 different DSM policies unsuccessfully implemented)

California, China, the UK and the USA have each successfully implemented and evaluated 9-10 different DSM policy categories, and the European Union (EU), India, Mexico and the USA have each unsuccessfully implemented 6-7 different DSM policy categories. In the case of the USA, its experience with DSM policy is the greatest of any country/state, which explains why it has both a high number of successful and unsuccessful policies. What is important here is the focus of this research on high quality evaluations. It is possible that some countries/states, such as the EU, may have implemented successful DSM policies (primarily through EU directives), but which have not been evaluated or have not been evaluated to a high quality standard, published in English and made publicly available online (as per the systematic review inclusion and exclusion criteria outlined in chapter three).

6.2 Key Contributions

The key contributions of this thesis to the energy policy field can be discussed in terms of empirical, methodological and conceptual contributions to research, which have useful implications for policy (discussed in section 6.3).

Empirically, the thesis provided the first systematic review of the DSM policy evaluation evidence base. The systematic review had a global coverage at both national (30 countries) and sub-national state-level (36 states) across six continents, and critically appraised the quality and global distribution of the evidence. Furthermore, the thesis examined twelve different categories of DSM policy in addition to nine DSM policy packages, providing one of the few demand-side analyses that have looked at DSM at its broadest (covering policies for energy efficiency, demand response and on-site generation and storage). However, the majority of evaluations focussed on energy efficiency policy.

The central part of the thesis was the identification of the key success and failure factors for each DSM policy category and for each country/state included in the sample. Under government resource constraints, the findings are useful for identifying where limited resources should be channelled in order to implement and evaluate a DSM policy successfully and to prevent policy failure. One of the central arguments of this thesis is that previous academic research has primarily focussed on the impacts of DSM policies rather than the reasons why DSM policies succeed or fail and the research has contributed towards filling this knowledge gap. Identifying where DSM policies have been successfully implemented (and where they have not) at both a national and state-level is an important contribution to the field.

Methodologically, the thesis further developed the theory and practice of using systematic reviews in energy policy research, particularly the practical application of the realist synthesis approach. Although the methodological approach was applied to DSM policy analysis, it can be readily adapted to the assessment of supply-side energy policies or environmental and climate policy more broadly. In addition, the approach developed and published a scale for critically appraising the quality of the evidence base, which could be applied to other aspects of energy, environmental and climate policy that focus on policy and programme mechanisms.

Furthermore, the research proposed and developed new data analysis techniques for the examination of policy mechanisms. The use of frequency, weighting and combined frequency-weighting analyses for determining the key factors for policy success and failure is a useful methodological contribution to the field. Another methodological contribution is the triangulation of MCDM analysis into systematic reviews in order to include a database of experts and to act as a form of cross-validation for the systematic review. Although MCDM analysis was a secondary method in the research, which was not analysed separately, the research developed an adapted form of MCDM method that had both quantitative and qualitative aspects to it. The method can also be applied to supply-side energy policy analysis as well as wider environmental and climate policy analysis.

Conceptually, the thesis proposed and published a new, holistic definition of DSM:

"*Demand-side management (DSM) refers to technologies, actions and programmes on the demand-side of energy meters that seek to manage or decrease energy consumption, in order to reduce total energy system expenditures or contribute to the achievement of policy objectives such as emissions reduction or balancing supply and demand*".

The definition was developed from a review of the definitions in the literature over the last thirty years (since the term 'demand-side management' was first coined in 1984 by Clark Gellings). The review collated the definitions and adapted them to be more relevant to policy objectives in the 21st century, such as the growing importance of carbon emissions reduction and energy security in the transition to low(er) carbon energy systems. Furthermore, the research made two additional theoretical contributions through the visualisation of the economics of policies to stimulate alternative utility business models to put demand-side options on an equal basis to supply-side options, and the visualisation of the 'perfect' DSM policy evaluation.

6.3 Policy Recommendations

The empirical, methodological and conceptual contributions to the field have important implications for policy. As the focus of the research is DSM policy, the findings are useful for governments around the world that are currently considering, designing or implementing demand-side policies. The key recommendations are summarised below by theme (policy mechanisms or policy analysis techniques) and by recommendation type (consideration, policy design, policy implementation or policy evaluation).

DSM policy mechanisms:

- *Consideration:* the absence of failure factors does not necessarily result in a policy succeeding
- *Consideration:* the absence of success factors does not necessarily result in a policy failing

- *Consideration:* the high quality documents identified in the systematic review should form the starting point for looking at the experiences of other countries/states – this reduces the governmental resources required to search for evidence on various DSM policies
- *Consideration:* the USA, the UK, California, France and China have the largest evidence bases for DSM policy and their experiences are useful for the future design and implementation of DSM policy
- *Policy design:* individual DSM policies are generally less resource intensive to implement and evaluate – however, DSM policies that tend to match well together in policy packages are: PS/IC, PS/LB/IC, IPBDR/PBDR, IC/L&S/MT, PS/LB, VP/L&S, PS/LB/UO/L&S, UBM/MT and IC/L&S
- *Policy implementation:* few DSM policies are likely to succeed without the required regulatory frameworks and appropriate incentives in place
- *Policy implementation:* DSM policies are likely to fail if there is a lack of monitoring throughout the policy period and technical issues through poorly administered programmes
- *Policy implementation:* under limited government resources, governments should focus their efforts on the specific success and failure factors by DSM policy and by country/state identified in the research

Policy analysis techniques:

- *Consideration:* systematic reviews should be used more widely in the development of energy policies – the research has provided practical methods to do this in relation to policy mechanisms for success and failure
- *Consideration:* under governmental time constraints, techniques for conducting rapid evidence assessments should be developed, which utilise the techniques developed in this research for systematic reviews – furthermore, different types of systematic review or rapid evidence assessment should be developed, such as those that focus on policy impacts
- *Policy evaluation:* it is important that DSM policies are appraised and evaluated using a combination of ex-ante and ex-post approaches

throughout the policy process (design, implementation and post-policy evaluation) – an adequate proportion of programme budgets should be dedicated to evaluation during the policy design stage

The UK is used as a case study in order to apply the above thesis statements and to provide policy recommendations. The UK has successfully implemented nine categories of DSM policy in the past, as shown below:

Successful DSM policies (9):

- Incentive payment-based demand response (IPBDR)
- Price-based demand response (PBDR)
- Infrastructure rollouts (IR)
- Utility obligations (UO)
- Performance standards (PS)
- Loans and subsidies (L&S)
- Alternative utility business models (UBM)
- Research and development programmes (R&D)
- Information campaigns (IC)

However, it has also unsuccessfully implemented four categories of DSM policy: market transformations (MT), labelling (LB), loans and subsidies (L&S) and the VP/L&S policy package (voluntary programmes combined with loans and subsidies). In addition to a continuation of the successfully implemented policies listed above, the UK can successfully implement MT, LB, L&S and VP/L&S if the following key success and failure factors are accounted for:

Market transformation (MT):

Success factors:

- Appropriate incentives

Failure factors:

- A lack of policy certainty
- A lack of monitoring
- Technical issues (primarily programme management issues)

Labelling (LB):

Success factors:

- Information infrastructure

Failure factors:

- Technical issues (primarily label design issues)

Loans and Subsidies (L&S):

Success factors:

- *No factors produced in the combined analysis*

Failure factors:

- *No factors produced in the combined analysis*

VP/L&S policy package:

Success factors:

- Appropriate incentives
- Consumer commitment
- Innovation
- Regulatory frameworks
- Long-term funding
- Cost-effectiveness
- Information infrastructure
- Industry engagement
- Legislative support
- Ease of implementation
- Policy continuity
- Wide coverage

Failure factors:

- A lack of sustainable funding
- Technical issues (primarily programme management issues)
- A lack of policy certainty
- Knowledge issues
- Inappropriate targeting
- A lack of consumer incentives
- Complexity

Furthermore, the UK could put a greater emphasis on policy packages, particularly the IC/L&S, UBM/MT and IPBDR/PBDR policy packages, which it has had less experience with to date. The following factors should be accounted for in order to increase the chance of policy success:

IC/L&S policy package:

Success factors:

- Clear definition of roles
- Industry engagement
- Ease of implementation
- Information infrastructure
- Regulatory frameworks
- Cost-effectiveness
- Innovation

Failure factors:

- Limited coordination between relevant parties
- A lack of sustainable funding

UBM/MT policy package:

Success factors:

- *No factors produced in the combined analysis*

Failure factors:

- *No factors produced in the combined analysis*

IPBDR/PBDR policy package:

Success factors:

- Regulatory frameworks
- Consumer commitment
- Legislative support
- Information infrastructure
- Physical infrastructure

Failure factors:

- A lack of sustainable funding
- Technical issues (primarily programme management issues)
- Poor consumer engagement

Case studies of policy recommendations, such as the above for the UK, are available for each of the 30 countries and 36 states and provinces included in the sample. This is due to the global focus of the research, as per the research questions, as opposed to taking a case study approach to the analysis.

6.4 Further Research

There are a number of areas for further research, which can be split into direct extensions to the thesis (sub-section 6.4.1) and wider research gaps (sub-section 6.4.2). In both cases, discussions cover both empirical and methodological areas for further research.

6.4.1 Extensions to the Thesis

With greater time and resources, the systematic review could be extended to include non-English documents (translators would be required), hand searching for non-electronic documents (access to specified libraries and organisations would be required) and the use of more search terms, such as the twelve additional search terms identified in sub-section 3.4.4. The use of hand searching and non-English documents would overcome issues of not being able to access older evaluations (from 1970s-1990s) that have not been digitised, as well as being able to extract relevant data and information from documents not written in English. This would also allow more detailed temporal analysis to be conducted, such as how the following aspects have changed over time: implementation patterns by DSM policy, implementation patterns by country/state, policy objectives for DSM, the analytical focus of evaluations and evidence quality.

A further area where the research could be extended is to explore the extent to which the policy impacts within the sample could be statistically aggregated and compared. However, it is likely that the contexts within which the policies were implemented are too diverse for statistical aggregation to produce useful or reliable results (as discussed in chapter three). Thus, a more appropriate extension to the research would be to increase the number of MCDM analysis interviews in the sample and to analyse the results separately from the systematic review, rather than triangulating them. This would allow for a greater exploration of not only the qualitative aspects of the MCDM analysis (which were used directly in the research), but also the quantitative aspects of the MCDM analysis (which were only used indirectly in the research). Greater resources would be needed to ensure that the sample was global and included experts outside of the UK and the east coast of the USA (where the 17 experts included in the research were located). Of particular priority would be to include experts based in California, France and China, as these countries/states have the largest evidence bases for DSM policy evaluation outside of the USA (at a national level) and the UK. The enlarged MCDM database would also act as a more extensive form of cross-validation for the systematic review.

6.4.2 Wider Research Gaps

Europe is one of the continents where the DSM policy evidence base is increasing. It would be interesting to update the systematic review post-2020 following the completion of important EU Directives, such as the *Energy Efficiency Directive* (*Directive 2012/27/EU*) and the *Smart Meter Rollout Directive* (*Directive 2009/72/EC*). The former includes a target to increase energy efficiency by 20% by 2020 and the latter requires the rollout of smart meters to at least 80% of small consumers by 2020. It is likely that the evaluation evidence base, particularly for utility obligations (due to the *Energy Efficiency Directive*) and infrastructure rollouts (due to the *Smart Meter Rollout Directive*), will increase as a result of these directives.

The thesis highlighted that the evidence base for South America and Africa is limited. A detailed evidence review of current and past experiences within these continents is a crucial area for further research, as they are much under-researched regions for DSM policy. As the published evidence is dominated by documents written in English by authors in English-speaking regions (such as in North America and Europe), it is likely that non-English evaluations conducted in South America (in Spanish or Portuguese) exist but have not been translated, and should be reviewed. This would be an important empirical contribution to the field. At the time of writing, the author proposed and is currently supervising an MSc dissertation at University College London (UCL) on the energy efficiency policy evidence base in South America using a rapid evidence assessment and the translation of documents written in Spanish (see Salinas Ivanenko, 2015). A similar research project for Africa is warranted.

An important methodological contribution to the field would be to further adapt the use of systematic reviews. Firstly, the same methods and techniques could be used to analyse supply-side energy policies or other areas of environmental and climate policy. Secondly, the same methods and techniques could be used to analyse DSM trials, pilots and non-government stimulated utility programmes rather than government policies. Thirdly, other types of systematic review, such as statistical meta-analysis and qualitative synthesis, should be explored to examine how (and if) they could be adapted to the analysis of energy and environmental policy. Linked to this, techniques should be developed to conduct rapid evidence assessments in the energy policy field, as the review method could be particularly useful for gathering evidence quickly under the time and resource constraints of most governments. Finally, the *Factor Weighting Scale* could be subjected to further methodological development by assessing potential differences in the evaluators' use of language and the frequency of discussion of specific words and phrases.

The thesis identified a number of DSM policies where the evaluation evidence base is limited in the literature. There are few high quality evaluations of large-scale research and development programmes, voluntary programmes and infrastructure rollouts. Methods for evaluating infrastructure rollouts will be particularly important to develop in the 2010s and early 2020s as a result of the *Smart Meter Rollout Directive*, as explained above. Furthermore, methods for evaluating DSM policy packages are currently limited in the literature and this is an important research gap that should be filled. However, this thesis argues that the priority area should be the development of methods to comprehensively evaluate information campaigns (IC), as this was one of the most frequently implemented categories of DSM policy in the research sample, but yet comprehensive IC policy evaluation methods were limited.

The research began to explore the area of policy transferability but due to time and resource constraints this could not be included in the final thesis. Thus, an important area for further research is to look at the transferability of successful DSM policies in particular countries/states to other countries/states. Of importance here is the level of policy transfer (either the direct copying of design and implementation processes or simply gaining inspiration and ideas) and the matching of countries/states with similar contexts, such as market structures, regulatory environments, cultures, climates, energy demands and energy system structures. Such research should also determine non-transferability, as identifying which countries/states' experiences should and should not be considered is useful to governments in the development of future DSM policies.

In conclusion, the main thesis statement is that policy mechanisms should be examined in addition to policy impacts in DSM policy evaluations, and DSM policies succeed where the required regulatory frameworks and appropriate incentives are in place and fail where there is a lack of monitoring for evaluation and technical issues with programme management.

7 Bibliography

7.1 Full References

1. Agnew, K., Burke, R. and Ham-su, P. (2009) "Participation of demand response resources in ISO New England's Ancillary Service Markets", *International Energy Program Evaluation Conference 2009*
2. Akobeng, A.K. (2005) "Understanding randomised controlled trials", *Archives of Disease in Childhood*, 90, pp. 840-844
3. Albadi and El-Saadany (2008) "A summary of demand response in electricity markets", *Electric Power Systems Research*, 78, pp. 1989-1996
4. Altrichter, H., Feldman, A., Posch, P. and Somekh, B. (2008) *Teachers investigate their work: an introduction to action research across the professions*, 2nd edition, Routledge, p. 147
5. American Recovery and Reinvestment Act of 2009, Washington DC: National Congress, USA
6. Anderson, D. (2006) "Power System Reserves and Costs with Intermittent Generation", Working Paper, UK Energy Research Centre (UKERC), UK
7. Arsenault, E., Bernard, J.-T. and Genest-Laplante, E. (1996) "Hydro-Québec energy savings programs: 'Watt' are they worth?", *Resource and Energy Economics*, 18 (1), pp. 65-81
8. Arteconi, A., Hewitt, N.J. and Polonara, F. (2012) "State of the art of thermal storage for demand-side management", *Applied Energy*, 93, pp. 371-389
9. Association for the Conservation of Energy (ACE) (2013) *Financing energy efficiency in buildings: an international review of best practice and innovation*, A report to the World Energy Council, October 2013
10. Atanasiu, B. and Constantinescu, T. (2011) "A comparative analysis of the energy performance certificates schemes within the European Union: Implementing options and policy recommendations", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)
11. Bachrach, D. (2003) "Energy Efficiency Leadership in California: Preventing the Next Crisis", *The Electricity Journal*, 16 (6), pp. 37-47
12. Bahgat, G. (2006) "Europe's energy security: challenges and opportunities", *International Affairs*, 82 (5), pp. 961-975
13. Banerjee, A. and Soloman, B.D. (2003) "Eco-labeling for energy efficiency and sustainability - a meta-evaluation of US programs", *Energy Policy*, 31, pp. 109-123
14. Bardach, E. (2005) *A practical guide for policy analysis: the eightfold path to more effective problem solving*, 2nd edition, New York: Chatham House Publishers
15. Barker, T., Ekins, P. and Foxon, T. (2007) "The macro-economic rebound effect and the UK economy", *Energy Policy*, 35, pp. 4935-4946
16. Barrett, M. (2006) *A renewable electricity system for the UK*, A response to the 2006 Energy Review, University College London (UCL), London, UK

17. Baxter, L.W. (1995) “Net Lost Revenue Adjustment (NLRA) Mechanisms for Utility DSM Programs”, *Energy*, 20 (12), pp. 1215-1223
18. Beaudin, M. *et al*. (2010) “Energy storage for mitigating the variability of renewable electricity sources: An updated review”, *Energy for Sustainable Development*, 14, pp. 302-314
19. Behavioural Insights Team (2011) *Annual Update 2010-2011 Report*, Cabinet Office, London, UK
20. Bennett, C.J. (1991a) "How States Utilize Foreign Evidence", *Journal of Public Policy*, 11, pp. 31-54
21. Bennett, C.J. (1991b) "What is Policy Convergence and What Cause it?", *British Journal of Political Science*, 21, pp. 215-233
22. Berger, V.W. (2006) “Is the Jadad score the proper evaluation of trials?”, The *Journal of Rheumatology*, 33 (8), pp. 1710-1712
23. Bertoldi, P. and Rezessy, S. (2007) "Assessment of White Certificate Schemes and their energy saving evaluation methods", *International Energy Program Evaluation Conference 2007*
24. Bertoldi, P. and Rezessy, S. (2008) "Tradable white certificate schemes: fundamental concepts", *Energy Efficiency*, 1, pp. 237-255
25. Betz, R., Jones, M.C., MacGill, I. and Passey, R. (2013) "Trading in energy efficiency in Australia: What are the lessons learnt so far?', *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)
26. Bin, S. and Jun, L. (2012) "Building Energy Efficiency Policies in China", *American Council for an Energy-Efficient Economy (ACEEE)*, Research Report E129, June 2012
27. Blumstein *et al*. (2000) "A theory-based approach to market transformation", *Energy Policy*, 28, pp. 137-144
28. Boait, P. (2009) *Energy services and ESCos – their benefits and implications for regulation and the consumer*, research report, De Montfort University, UK
29. Boza-Kiss, B., Moles-Grueso, S. and Ürge-Vorsatz, D. (2013) “Evaluating policy instruments to foster energy efficiency for the sustainable transformation of buildings”, *Opinion in Environmental Sustainability*, 5, pp. 163-176
30. Brazil National Law 9.991/2000 on energy efficiency (PEE), ANEEL, Brazil
31. Broc, J.-S., Melo, C.A. and Jannuzzi, G. (2012) "Detailed comparison of Brazilian and French obligation schemes to promote energy efficiency", *International Energy Program Evaluation Conference 2012*
32. Brown, B. (2014) *A meta-evaluation of energy efficiency evaluations*, PhD thesis, Western Michigan University, USA
33. Brown, M.A. and Mihlmester, P.E. (1995) "What has DSM achieved in California?", report number CONF-950643-1, Department of Energy (DoE), USA
34. Bukarica, V., Debrecin, N., Borkovic, Z.H. and Pesut, D. (2012) "Evaluating energy savings arising from NEEAP implementation: lessons learned in Croatia", *International Energy Program Evaluation Conference 2012*
35. Bundgaard, S.S., Dyhr-Mikkelsen, K., Larsen, A.E. and Togeby, M. (2013a) "Energy Efficiency Obligation Schemes in the EU - Lessons Learned from Denmark", First quarter 2013, *International Association for Energy Economics (IAEE)*

36. Bundgaard, S.S., Togeby, M., Dyhr-Mikkelsen, K., Sommer, T., Kajaerbye, V.H. and Larsen, A.E. (2013b) "Spending to save: evaluation of the energy efficiency obligation in Denmark", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)
37. Bunn, D.W. and Seigal, J.P. (1983) "Television peaks in electricity demand", *Energy Economics*, January
38. Cabeza, L.F., Ürge-Vorsatz, D., McNeil, M.A., Barreneche, C. and Serrano, S. (2014) "Investigating greenhouse challenge from growing trends of electricity consumption through home appliances in buildings", *Renewable and Sustainable Energy Reviews*, 36, pp. 188-193
39. Campbell Collaboration website: http://www.campbellcollaboration.org
40. Capgemini (2008) *Demand Response: a decisive breakthrough for Europe*, in collaboration with Vaasaett and Enerdata
41. Cappers, P., Goldman, C. and Kathan, D. (2010) "Demand response in U.S. electricity markets: Empirical evidence", *Energy*, 35 (4), pp. 1526-1535
42. Cappers, P., Goldman, C.A. and Kathan, D. (2009) "Demand Response in US Electricity Markets: Empirical Evidence", report number LBNL-2124E, Ernest Orlando Lawrence Berkeley National Laboratory, USA
43. Carbon Reduction Commitment Energy Efficiency Scheme, 2010, Department of Energy and Climate Change (DECC), UK
44. Centre for Strategy & Evaluation Services (2012) "Evaluation of the Ecodesign Directive", Final report, UK
45. Chatterton, T. (2011) *An introduction to thinking about 'energy behaviour': a multi-model approach*, Paper to DECC, London, UK
46. Cheng, C-C. (2005) *Electricity demand-side management for an energy efficient future in China: technology options and policy priorities*, PhD Thesis, Massachusetts Institute of Technology, USA, pp. 43
47. Chinnow, J., Bsufka, K., Schmidt, A.-D., Bye, R., Camtepe, A. and Albayrak, S. (2011) "A simulation framework for smart meter security evaluation", *2011 IEEE International Conference on Smart Measurements for Future Grids (SMFG)*, pp. 1-9, 14-16th November 2011
48. Chitnis, M., Sorrell, S., Druckman, A., Firth, S.K. and Jackson, T. (2013) "Turning lights into flights: estimating direct and indirect rebound effects for UK households", *Energy Policy*, 55, pp. 234-250
49. Clark, H.D., Wells, G.A., Huët, C., McAlister, F.A., Salmi, L.R., Fergusson, D. and Laupacis, A. (1999) "Assessing the quality of randomised trials: reliability of the Jadad scale", *Controlled Clinical Trials*, 20 (5), pp. 448-452
50. Clastres, C. (2011) "Smart grids: Another step towards competition, energy security and climate change objectives", *Energy Policy*, 39 (9), pp. 5399-5408
51. Climate Change Act of 2008, UK
52. Climate Change Agreements (CCAs), 2001, Department of Trade and Industry (DTI), UK
53. Climate Change Levy (CCL), 2001, Department of Trade and Industry (DTI), UK
54. Cochrane Collaboration website: http://www.cochrane.org/

55. Colby, J. and Davis, S. (2011) "Can short term ARRA stimulus funding achieve long term market transformation?", *International Energy Program Evaluation Conference 2011*
56. Collaboration for Environmental Evidence (CEE) website: http://www.environmentalevidence.org
57. Combined Heat and Power Association (CHPA) (2014) “What is CHP?”, accessed on 11/11/14, website article can be downloaded from: http://www.chpa.co.uk/what-is-chp_15.html
58. Congress.gov (2015) *Energy Efficiency Improvement Act of 2015*, S.128 – 114th Congress (2015-2016), accessed on 24/02/15, can be downloaded from: https://www.congress.gov/bill/114th-congress/senate-bill/128?q=%7B%22search%2 2%3A%5B%22Energy+Efficiency+Improvement+Act%22%5D%7D
59. Connor, P.M., Baker, P.E., Xenias, D., Balta-Ozkan, N., Axon, C.J. and Cipcigan, L. (2014) “Policy and regulation for smart grids in the United Kingdom”, *Renewable and Sustainable Energy Reviews*, 40, pp. 269-286
60. Cooke, D. (2011) *Empowering Customer Choice in Electricity Markets*, Information Paper, October 2011, International Energy Agency (IEA)
61. Crossley, D. (2008) "Tradeable energy efficiency certificates in Australia", *Energy Efficiency*, 1, pp. 267-281
62. Crossley, D. (2010) "International best practice in using energy efficiency and demand management to support electricity networks", *Report 4 - Scaling the Peaks: Demand Management and Electricity Networks*, Australian Alliance to Save Energy, Australia
63. Crossley, D. (2011) “Tempo Electricity Tariff – France”, 11th case study article of Task XV – Network-Driven DSM, *IEA DSM Programme*
64. Crossley, D., Maloney, M. and Watt, G. (2000) “Developing mechanisms for promoting demand-side management and energy efficiency in changing electricity businesses”, *IEA DSM Programme*, Task VI Research Report No. 3, Energy Futures Australia, August 2000
65. Cuijper, C. and Koops, B.J. (2012) “Smart metering and privacy in Europe – lessons from the Dutch case”, in: *European Data Protection: Coming of Age*, Dordrecht: Springer, pp. 269-293
66. Daly *et al*. (2007) “A hierarchy of evidence for assessing qualitative health research”, *Journal of Clinical Epidemiology*, 60, pp. 43-49
67. Daramola, O.O. and Rhee, J.S. (2011) “Rating evidence in medical literature”, *American Medical Association Journal of Ethics*, 13 (1), pp. 46-51
68. Darby, S. (2006) *The effectiveness of feedback on energy consumption – a review for DEFRA*, Environmental Change Institute, University of Oxford
69. Davito, B., Tai, H. and Uhlaner, R. (2010), *The Smart Grid and the Promise of Demand-Side Management*, Report, McKinsey and Co.
70. Dawnay and Shah (2011) “Behavioural economics: seven key principles for environmental policy”, Chapter 4 (pp. 74-99), in: Dietz, S., Mitchie, J. and Oughton, C. (2011) *The Political Economy of the Environment: An Interdisciplinary Approach*, Routledge, UK
71. Deane, J.P., Ó Gallachóir, B.P. and McKeogh, E.J. (2010) “Techno-economic review of existing and new pumped hydro energy storage plant”, *Renewable and Sustainable Energy Reviews*, 14, pp. 1293-1302
72. Decent Homes Standard, 2000, Department of Environment, Food and Rural Affairs (DEFRA), UK

73. Denzin, N. (2006) *Sociological Methods: A Sourcebook*, 5th edition, Aldine Transaction, ISBN 978-0-202-30840-1
74. Depuru, S.S.S.R., Wang, L. and Devabhaktuni, V. (2011) “Smart meters for power grid: Challenges, issues, advantages and status”, *Renewable & Sustainable Energy Reviews*, 15 (6), pp. 2736-2742
75. Devine-Wright, H. and Devine-Wright, P. (2004) “From demand side management to demand side participation: tracing an environmental psychology of sustainable electricity system evolution”, *Journal of Applied Psychology*, 6 (3-4), pp. 167-177
76. Di Santo, D., Forni, D., Venyurini, V. and Biele, E. (2011) "The white certificate scheme: the Italian experience and proposals for improvement", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)
77. Didden, M.H. and D’haeseleer, W.D. (2003) “Demand Side Management in a competitive European market: who should be responsible for its implementation?” *Energy Policy*, 31, pp. 1307-1314
78. DiSanto, D. (2014) “White certificates in Italy”, International Energy Policies and Programmes Evaluation Conference, 8th September, pre-conference workshop presentation, Federazione Italiana per l'uso Razionale dell'Energia (FIRE), Berlin, Germany
79. Divya, K.C. and Østergaard, J. (2009) “Batter energy storage technology for power system – an overview”, *Electric Power Systems Research*, 79 (4), pp. 511-520
80. Dixon-Woods, M., Agarwal, S., Jones, D., Young, B. and Sutton, A. (2005) “Synthesising qualitative and quantitative evidence – a review of possible methods”, *Journal of Health Services Research Policy*, 10 (1), pp. 45-53
81. Dockerty, T., Dockerty, T., Lovett, A., Papathanasopoulou, E., Beaumont, N., Wang, S. and Smith, P. (2009) “UKERC Energy Strategy Under Uncertainties: Interactions between the Energy System, Ecosystem Services and Natural Capital”, Working Paper, UKERC/WP/FG/2014/010, *UK Energy Research Centre (UKERC)*, UK
82. Doris, E., Cochran, J. and Vorum, M. (2009) "Energy Efficiency Policy in the United States: Overview of Trends at Different Levels of Government", report number NREL/TP-6A2-46532, National Renewable Energy Laboratory, Department of Energy (DoE), USA
83. Doris, E., Cochran, J. and Vorum, M. (2009) “Energy efficiency policy in the United states: overview of trends at different levels of government”, *National Renewable Energy Laboratory*, Technical report: NREL/TP-6A2-46532, USA
84. Drew, T. (2009) "An assessment of California's energy efficiency incentive mechanism", *International Energy Program Evaluation Conference 2009*
85. Droste-Franke, B., Paal, B.P., Rehtanz, C., Sauer, D.U., Schneider, J-P., Schreurs, M. and Ziesemer, T. (2012) *Balancing renewable electricity: energy storage, demand side management, and network extension from an interdisciplinary perspective*, Berlin: Springer
86. Druckman, A., Chitnis, M., Sorrell, S. and Jackson, T. (2011) “Missing carbon reductions? Exploring rebound and backfire effects in UK households”, *Energy Policy*, 39, pp. 3572-3581

87. Effendi, P. and Courvisanos, J. (2012) "Political aspects of innovation: examining renewable energy in Australia", *Renewable Energy*, 38 (1), pp. 245-252
88. Eissa, M.M. (2011) "Demand side management program evaluation based on industrial and commercial field data", *Energy Policy*, 39, pp. 5961-5969
89. Ekins, P. (p.ekins@ucl.ac.uk), 20th February 2015, Comment on policy evaluation, E-mail to P. Warren (peter.warren.10@ucl.ac.uk)
90. El Bakari, K., Myrzik, J.M.A. and Kling, W.L. (2009) "Prospects of a virtual power plant to control a cluster of distributed generation and renewable energy sources", *2009 Proceedings of the 44th International Universities Power Engineering Conference (UPEC),* pp. 1-5, 1-4th September 2009, Glasgow
91. Energy Act of 2004, London: HMSO, UK
92. Energy Act of 2008, London: HMSO, UK
93. Energy Act of 2011, London: HMSO, UK
94. Energy Bill of 2011, Department of Energy and Climate Change (DECC), UK
95. Energy Company Obligation 1 (ECO 1), 2013, Department of Energy and Climate Change (DECC), UK
96. Energy Company Obligation 2 (ECO 2), 2015, Department of Energy and Climate Change (DECC), UK
97. Energy Efficiency Commitment 1 (EEC 1), 2002, Department of Environment, Food and Rural Affairs (DEFRA), UK
98. Energy Efficiency Commitment 2 (EEC 2), 2005, Department of Environment, Food and Rural Affairs (DEFRA), UK
99. Energy Efficiency Standards of Performance 1 (EESoP 1), 1994, Department of Environment, Food and Rural Affairs (DEFRA), UK
100. Energy Efficiency Standards of Performance 2 (EESoP 2), 1998, Department of Environment, Food and Rural Affairs (DEFRA), UK
101. Energy Efficiency Standards of Performance 3 (EESoP 3), 2000, Department of Environment, Food and Rural Affairs (DEFRA), UK
102. Energy Independence and Security Act of 2007, Washington DC: National Congress, USA
103. Energy Information Administration (EIA) (2011) *Country Reports*, can be downloaded from: http://www.eia.gov/countries/
104. Energy Policy Act of 1992, Washington DC: National Congress, USA
105. Energy Policy Act of 2005, Washington DC: National Congress, USA
106. Energy Policy and Conservation Act of 1975, Washington DC: National Congress, USA
107. Eto, J.H. (1994) "The theory and practice of decoupling", *eScholarship*, Lawrence Berkeley Laboratory, USA
108. European Commission (2008) *EU Climate and Energy Package*, Brussels, Belgium
109. European Commission (2012) "Making the internal energy market work", *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions*, COM(2012) 663 final, 15/11/2012, Brussels, Belgium

110. European Council Directive 2002/91/EC on energy performance of buildings
111. European Council Directive 2009/72/EC on smart meter rollout
112. European Council Directive 92/42/EEC on improving boiler efficiency
113. European Council Directive 92/75/EEC on energy labelling of products
114. European Council Directive 93/76/EEC on improving energy efficiency
115. European Council Directive 96/92/EC on common rules for the internal electricity market
116. European Council Directive 98/30/EC on common rules for the internal gas market
117. European Smart Grids Technology Platform (2006) *Vision and Strategy for Europe's Electricity Networks of the Future*, European Commission, EUR 22040, Brussels, Belgium
118. European Union (EU), *Treaty on European Union (Consolidated Version), Treaty of Maastricht*, 7 February 1992, Official Journal of the European Communities C 325/5; 24 December 2002, available at: http://www.unhcr.org/refworld/docid/3ae6b39218.html, accessed 14/03/12
119. European Union Directive 2010/30/EU on energy labelling of products
120. European Union Directive 2010/31/EU on energy performance of buildings
121. European Union Directive 2012/27/EU on energy efficiency
122. Evans, A., Strezov, V. and Evans, T.J. (2012) “Assessment of utility energy storage options for increased renewable energy penetration”, *Renewable and Sustainable Energy Reviews*, 16, pp. 4141-4147
123. Eyre, N. (2008) "Regulation of energy suppliers to save energy - lessons from the UK debate", *British Institute of Energy Economists (BIEE)*
124. Eyre, N., Pavan, M. and Bodineau, L. (2009) "Energy company obligations to save energy in Italy, the UK and France: what have we learnt?", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)
125. Feed-In Tariffs, 2010, Department of Energy and Climate Change (DECC), UK
126. Fischer, F. (1995) *Evaluating public policy*, Chicago: Nelson-Hall Publishers
127. Fleiter, T., Gruber, E., Eichhammer, W. and Worrell, E. (2012) "The German energy audit program for firms – a cost-effective way to improve energy efficiency?", *Energy Efficiency*, 5, pp. 447-469
128. Friedmann, R. and Sheinbaum, C. (1998) "Mexican electric end-use efficiency: experiences to date", *Annual Review of Energy and the Environment*, 23, pp. 225-252
129. Galarraga, I., Gonzàlez-Eguino, M. and Markandya, A. (2011) *Handbook of Sustainable Energy*, Cheltenham: Edward Elgar Publishing Limited, p. 5

130. Ganann, R., Ciliska, D. and Thomas, H. (2010) “Expediting systematic reviews: methods and implications of rapid reviews”, *Implementation Science*, 5 (56)

131. Geller, H. and Nadel, S. (1994) “Market transformation strategies to promote end-use efficiency”, *Annual Review of Energy and the Environment*, 19, pp. 301-346

132. Gellings, C. W. (1996) "Then and now: The perspective of the man who coined the term 'DSM'", *Energy Policy*, 24 (4), pp. 285-288

133. Gellings, C. W. (2000) "Before demand-side management is discarded, let's see what pieces should be kept", *OPEC Review*, 24 (1), pp. 61-70

134. Gellings, C.W. (1985) “The concept of demand-side management for electric utilities”, *Proceedings of the IEEE*, 73 (10)

135. Gellings, C.W. and Chamberlin, J.H. (1993) *Demand-Side Management: Concepts and Methods*, 2nd Edition, The Fairmont Press, Inc., USA

136. Gilleo, A., Chittum, A., Farley, K., Neubauer, M., Nowak, S., Ribeiro, D. and Vaidyanathan, S. (2014) “The 2014 State Energy Efficiency Scorecard”, Research Report U1408, *American Council for an Energy-Efficient Economy (ACEEE)*, October 2014

137. Gillich, A. and Sunikka-Blank, M. (2013) "Barriers to domestic energy efficiency – an evaluation of retrofit policies and market transformation strategies", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)

138. Gillingham, K., Newell, R. and Palmer, K. (2006) "Energy efficiency policies: a retrospective examination", *Annual Review of Environment and Resources*, 31, pp. 161-192

139. Gillingham, K., Newell, R.G. and Palmer, K. (2009) “Energy efficiency economics and policy”, *Annual Review of Resource Economics*, NBER Working Paper Series, National Bureau of Economic Research

140. Giraudet, L.-G., Bodineau, L. and Finon, D. (2012) "The costs and benefits of white certificates schemes", *Energy Efficiency*, 5, pp. 179-199

141. Gold, R. and Nadel, S. (2011a) "Assessing the Harvest: Implementation of the Energy Efficiency Provisions in the Energy Policy Act of 2005", *American Council for an Energy-Efficient Economy (ACEEE)*, Research Report E113, March 2011

142. Gold, R. and Nadel, S. (2011b) "Energy Efficiency Tax Incentives, 2005-2011: How Have They Performed?", *American Council for an Energy-Efficient Economy (ACEEE)*, White Paper on Tax Incentives, June 2011

143. Gold, R., Nadel, S., Laitner, J.A. and deLaski, A. (2011) "Appliance and Equipment Efficiency Standards: A Moneymaker and Job Creator", *American Council for an Energy-Efficient Economy (ACEEE)*, Research Report A111, January 2011

144. Goldman, C.A. and Kito, M.S. (1994) "Review of Demand-Side Bidding Programs: Impacts, Costs and Cost-Effectiveness", report number LBL-35021, Lawrence Berkeley National Laboratory and the University of California, USA

145. Goldman, C.A., Stuart, E., Hoffman, I., Fuller, M.C. and Billingsley, M.A. (2011) "Interactions between Energy Efficiency Programs funded under the Recovery Act and Utility Customer-Funded Energy Efficiency

Programs", report number LBNL-4322E, Ernest Orlando Lawrence Berkeley National Laboratory, USA

146. Gough, D., Oliver, S. and Thomas, J. (2012) *An introduction to Systematic Reviews*, SAGE Publications Ltd., London, UK

147. Granovetter, M.S. (1973) “The strength of weak ties”, *The American Journal of Sociology*, 78 (6), pp. 1360-1380

148. Greenhouse Gas Abatement Scheme (GGAS), New South Wales, Australia

149. Greening, L.A. (2010) “Demand response resources: who is responsible for implementation in a deregulated market?”, *Energy*, 35, pp. 1518-1525

150. Grubb, M. (2014) “Tried and tested – four decades of energy efficiency policy”, chapter five in: Grubb (2014) *Planetary Economics: Energy, climate change and the three domains of sustainable development*, Routledge, UK

151. Gruber, E., Fleiter, T., Mai, M. and Frahm, B.-J. (2011) "Efficiency of an energy audit programme for SMEs in Germany – results of an evaluation study", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)

152. Guertler, P., Royston, S. and Robson, D. (2013) "Somewhere between a ‘Comedy of errors’ and ‘As you like it’? A brief history of Britain’s ‘Green Deal’ so far", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)

153. Gurwick, N.P., Moore, L.A., Kelly, C. and Elias, P. (2013) “A systematic review of biochar research, with a focus on its stability in situ and its promise as a climate mitigation strategy”, *PLoS ONE*, 8 (9)

154. Haas, R., Nakicenovic, N., Ajanovic, A., Faber, T., Kranzl, L., Müller, A. and Resch, G. (2008) “Towards sustainability of energy systems: a primer on how to apply the concept of energy services to identify necessary trends and policies”, *Energy Policy*, 36 (11), pp. 4012-4021

155. Hadley, S. and Hirst, E. (1995) "Utility DSM programs from 1989 through 1998: continuation or cross roads?", report number ORNL/CON-405, Oak Ridge National Laboratory, USA

156. Hamidi, V., Li, F. and Robinson, F. (2009) “Demand Response in the UK’s domestic sector”, *Electric Power Systems Research*, 79 (12), pp. 1722-1726

157. Hamilton, B., Plunkett, J. and Wickenden, M. (2002) “Gauging success of the nation’s first efficiency utility: Efficiency Vermont’s first two years”, *2002 ACEEE Summer Study on Energy Efficiency in Buildings – Proceedings*, American Council for an Energy-Efficient Economy (ACEEE), USA

158. Hamilton, J.D. (2011) *Historical Oil Shocks*, NBER Working Paper Series, Working Paper 16790, National Bureau of Economic Research, USA, http://www.nber.org/papers/w16790

159. Harden, A. and Thomas, J. (2005) “Methodological issues in combining diverse study types in Systematic Reviews”, *International Journal of Social Research Methodology*, 8 (3), pp. 257-271

160. Harmelink, M., Nilsson, L. and Harmsen, R. (2008) "Theory-based policy evaluation of 20 energy efficiency instruments", *Energy Efficiency*, 1, pp. 131-148

161. Harrington, L. and Damnics, M. (2004) "Energy labelling and standards programs throughout the world", *The National Appliance and Equipment Energy Efficiency Committee* (NAEEEC), NAEEEC Report 2004/04, Australia
162. Hart, C. (2002) *Doing a Literature Search*, London: SAGE Publications Ltd.
163. Hatchwell, P. (2013) "Smart meter roll-out complications force delay", *The ENDS Report*, Environmental Data Services, accessed on 09/10/14, can be downloaded at: http://www.endsreport.com/38905/smart-meter-roll-out-complications-force-delay
164. Hatchwell, P. (2014a) "MPs savage 'failing' Green Deal 18 months on", *The ENDS Report*, Environmental Data Services, accessed on 13/10/14, can be downloaded at: http://www.endsreport.com/45393/mps-savage-failing-green-deal-18-months-on
165. Hatchwell, P. (2014b) "Key Home Improvement Fund closes again", *The ENDS Report*, Environmental Data Services, accessed on 19/12/14, can be downloaded at: http://www.endsreport.com/46595/key-home-improvement-fund-budget-closes-again
166. Hayes, S., Nadel, S., Kushler, M. and York, D. (2011) "Carrots for Utilities: Providing Financial Returns for Utility Investments in Energy Efficiency", Report Number U111, American Council for an Energy-Efficient Economy (ACEEE), Washington D.C., USA
167. Heffner, G.C., Lees, E., Crossley, D. and Migden-Ostrander, J. (2012) "Evaluating policies for energy provider-delivered energy efficiency", *International Energy Program Evaluation Conference 2012*
168. Hill, M. (2009) *The Public Policy Process*, 5th Edition, Harlow: Pearson Education Limited, p. 143
169. Hills, J. (2012) "Getting the measure of fuel poverty – final report of the Fuel Poverty Review", CASE report 72, *Centre for Analysis of Social Exclusion (CASE)*, March 2012
170. Hirst, E. (2002) "The financial and physical insurance benefits of price-responsive demand", *The Electricity Journal*, pp. 66-73
171. Hoenkamp, R., Huitema, G.B. and de Moor-van Vugt, A.J.C. (2011) "The neglected consumer: the case of the smart meter rollout in the Netherlands", *Renewable Energy Law and Policy Review*, pp. 269-282
172. Hogwood, B.W. and Gunn, L.A. (1984) *Policy Analysis for the Real World*, Oxford University Press
173. Home Energy Conservation Act of 1996, London: HMSO, UK
174. Hong, T. (2009) "A close look at the China Design Standard for Energy Efficiency of Public Buildings", *Energy and Buildings*, 41, pp. 426-435
175. Horowitz (2013) "Purchased energy and policy impacts in the US manufacturing sector", *Energy Efficiency*, April 2013
176. Hsiao, C. (2003) *Analysis of Panel Data*, 2nd edition, New York: Cambridge University Press
177. ICF International (2007) *Introduction to energy performance contracting*, report prepared for the US Environmental Protection Agency (EPA), USA
178. IEA DSM Programme (2004) "INDEEP Analysis Report 2004", *Task 1 Subtask 8 - International Database on Demand-Side*

Management Technologies and Programmes, International Energy Agency (IEA)
179. IEA DSM Programme (2005) "Evaluation Guidebook Volume 2 - Country Reports and Case Examples Used for the Evaluation Guidebook", *Task 1 Subtask 9 - Evaluation Guidebook on the impact of DSM and Energy Efficiency Programmes for Kyoto's GHG Targets*, International Energy Agency (IEA)
180. IEA DSM Programme (2006) "Task 14 Final Report", *Task 14 - Market Mechanisms for White Certificate Trading*, International Energy Agency (IEA)
181. IEA DSM Programme (2008) "Evaluation and Acquisition of Network-driven DSM Resources", *Task 15 - Network driven DSM*, International Energy Agency (IEA)
182. IEA DSM Programme (2012) “Best practices in designing and implementing energy efficiency obligation schemes”, Research report prepared by the Regulatory Assistance Project (RAP), Task XXII of the IEA DSM Programme, June 2012
183. International Energy Agency (IEA) (2009) *Gadgets and Gigawatts – Policies for Energy Efficient Electronics*, Paris, France
184. International Energy Agency (IEA) (2011) *Country Statistics*, can be downloaded from: http://www.iea.org/statistics/
185. International Energy Agency (IEA) (2011) *Energy Policies of IEA Countries Reviews*, can be downloaded from: http://www.iea.org/publications/countryreviews/
186. International Energy Agency (IEA) (2012) *Best practices in designing and implementing energy efficiency obligation schemes*, Research Report: Task XXII of the IEA DSM Programme, prepared by the Regulatory Assistance Project (RAP)
187. International Energy Agency (IEA) (2014) “Energy efficiency”, accessed on 10/11/14, website article can be downloaded from: http://www.iea.org/topics/energyefficiency/
188. International Institute for Energy Conservation (IIEC) *Demand Side Management: Best Practices Guidebook for Pacific Island Power Utilities*, Report prepared for the South Pacific Applied Geoscience Commission (SOPAC) and the United Nations Department of Economic and Social Affairs (UNDESA)
189. International Performance Measurement and Verification Protocol (IPMVP) website, accessed on 02/01/15, can be downloaded from: http://www.evo-world.org/index.php?lang=en
190. International Standards Organisation (ISO) (2014) *Information on ISO 500001 and ISO 140001*, accessed 24/11/14, can be downloaded from: http://www.iso.org/iso/home.htm
191. Jadad, A. (1998) *Randomised controlled trials: a users guide*, London: BMJ Books, UK
192. John, P. (1998) *Analysing Public Policy*, Creative Print and Design (Wales), UK, pp. 196
193. John, P. (2011) *Making Policy Work*, Routledge, UK, pp. 10
194. Johnson-Brown, J. and Reingen, P.H. (1987) “Social ties and word-of-mouth referral behaviour”, *Journal of Consumer Research*, 14
195. Joskow, P.L. (2008) “Lessons learned from electricity market liberalisation”, *The Energy Journal*, 10, pp. 9-42

196. Kayikci, M.S. (2011) “The European Third Energy Package: how significant for the liberalisation of energy markets in the European Union?”, available at *Social Science Research Network* (SSRN): http://ssrn.com/abstract=2102161

197. Kellstedt, P.M. and Whitten, G.D. (2013) *The Fundamentals of Political Science Research*, 2nd edition, New York: Cambridge University Press

198. Kerr, D., Lemaire, X. and Owen, G. (2011) “An annotated bibliography and reference guide on energy efficiency and demand side management”, *SERN Literature Review 2011*, University College London (UCL), London, UK

199. Kim, J-H. and Shcherbakova, A. (2011) “Common failures of demand response”, *Energy*, 36, pp. 873-880

200. Kimura, O. (2009) "Is public R&D in energy efficiency really effective? – a case in Japan and its implications", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)

201. Kimura, O. (2010) “Japanese Top Runner approach for energy efficiency standards”, *SERC Discussion Paper*, SERC09035, Central Research Institute of Electric Power Industry

202. Kunseler, E. (2007) *Towards a new paradigm of science in scientific policy advising*, Environment and Health Department, NUSAP.net, p. 5, can be downloaded from: http://www.nusap.net/downloads/KunselerEssay2007.pdf

203. Kushler, M. and Vine, E. (2003) "Examining California's Energy Efficiency Policy Response to the 2000/2001 Electricity Crisis: Practical Lessons Learned Regarding Policies, Administration, and Implementation", *American Council for an Energy-Efficient Economy (ACEEE)*, Research Report U033, March 2003

204. Kushler, M., Nowak, S. and Witte, P. (2012) "A National Survey of State Policies and Practices for the Evaluation of Ratepayer-Funded Energy Efficiency Programs", *American Council for an Energy-Efficient Economy (ACEEE)*, Research Report U122, February 2012

205. Kushler, M., York, D. and Witte, P. (2004) "Five Years In: An Examination of the First Half-Decade of Public Benefits Energy Efficiency Policies", *American Council for an Energy-Efficient Economy (ACEEE)*, Research Report U041, April 2004

206. Lee, A.D., Onisko, S.A., Sandahl, L.J. and Butler, J. (1994) "Everyone wins! — A program to upgrade energy efficiency in manufactured housing", *The Electricity Journal*, 7 (2), pp. 77-87

207. Lee, W.L. and Yik, F.W.H. (2004) “Regulatory and voluntary approaches for enhancing building energy efficiency”, *Progress in Energy and Combustion Science*, 30 (5), pp. 477-499

208. Lees, E. (2006) "Evaluation of the Energy Efficiency Commitment 2002-2005", Report to DEFRA, prepared by Eoin Lees Energy, 28th February 2006, UK

209. Lees, E. (2008) "Evaluation of the Energy Efficiency Commitment 2005-2008", Report to DEFRA, prepared by Eoin Lees Energy, 28th February 2006, UK

210. Leopold, L.B., Clarke, F.E., Hanshaw, B.B. and Balsley, J.R. (1971) *A procedure for evaluating environmental impact*, Washington: Geological Survey Circular 645, p. 13

211. Lepetitgaland, K.K., Vasconcellos, L.E.M., Magalhaes, L.P., Costa, A.L.P., Costa, L.F.T. and Friedmann, R. (2011) "The strategic value of evaluation for Brazil and Neighbouring countries: the experience of Procel", *International Energy Program Evaluation Conference 2011*
212. Likert, R. (1932) "A Technique for the Measurement of Attitudes", *Archives of Psychology*, 140, pp. 1–55
213. Lise, W. and Kruseman, G. (2008) “Long-term price and environmental effects in a liberalized electricity market”, *Energy Economics*, 30 (2), pp. 230-248
214. Lovins, A.B. (1990) “The negawatts revolution”, *The Conference Board Magasine*, 27 (9), pp. 18-23
215. Low Carbon Innovation Programme, 2001, Department of Trade and Industry (DTI), UK
216. Lowry, M., Irwin, S. and Waeckerlin, E. (2004) "Demand Response in the West: Lessons for States and Provinces", report prepared for the Department of Energy (DoE), USA
217. Mallaburn, P.S. and Eyre, N. (2013) "Lessons from energy efficiency policy and programmes in the UK from 1973 to 2013", *Energy Efficiency*, February 2013
218. Marques, V., Bento, N. and Costa, P.M. (2014) “The “Smart Paradox” - Stimulate the deployment of smart grids with effective regulatory instruments”, *Energy*, 69, pp. 96-103
219. Mau, P., Eyzaguirre, J., Jaccard, M., Collins-Dodd, C. and Tiedemann, K. (2008) “The ‘neighbour effect’: simulating dynamics in consumer preferences for new vehicle technologies”, *Ecological Economics*, 68, pp. 504-516
220. McClone, C. (2014a) “Overwhelmed Green Deal fund frozen”, *The ENDS Report*, 475, September 2014, pp. 29, Environmental Data Services, accessed on 07/10/14, can be downloaded at: http://www.endsreport.com/45161/overwhelmed-green-deal-fund-frozen
221. McClone, C. (2014b) “Green Deal Home Improvement Fund extension granted”, *The ENDS Report*, Environmental Data Services, accessed on 07/10/14, can be downloaded at: http://www.endsreport.com/45658?DCMP=EMC-ENDSRPTENERGYBULLETIN
222. McClone, C. (2014c) “GDF Suez, Scottish Power and SSE fined for missing environmental targets”, *The ENDS Report*, Environmental Data Services, accessed on 22/12/14, can be downloaded from: http://www.endsreport.com/46578/gdf-suez-scottish-power-and-sse-fined-for-missing-environmental-targets
223. McConnell, A. (2010) “Policy success, policy failure and grey areas in-between”, *Journal of Public Policy*, 30, pp. 345-362
224. McDowall, W., Anandarajah, G., Dodds, P. and Tomei, J. (2012) “Implications of sustainability constraints on UK bioenergy development: assessing optimistic and precautionary approaches with UK MARKAL”, *Energy Policy*, 47, pp. 424-436
225. McKenna, E., Richardson, I. and Thomson, M. (2012) “Smart meter data: balancing consumer privacy concerns with legitimate applications”, *Energy Policy*, 41, pp. 807-814
226. McNerney, R.A. (1998) Changing Structure of the Electric Power Industry: An Update, Chapter 4, pp. 27-28, DIANE Publishing

227. McRae, M., James, A., Kim, A. and Spahic, M. (2011) "Measuring diffusion in a market transformation program", *International Energy Program Evaluation Conference 2011*

228. Meyers, E.M. and Hu, G.M. (1999) "Demand-Side Carbon Reduction Strategies in an Era of Electric Industry Competition", *The Electricity Journal*, 12 (1), pp. 72-81

229. Miles, M.B. and Huberman, M. (1994) *Qualitative data analysis: an expanded sourcebook*, London: Sage

230. Ming, Z., Song, X., Mingjuan, M., Lingyun, L., Min, C. and Yuejin, W. (2013) “Historical review of demand side management in China: Management content, operation mode, results assessment and relative incentives”, *Renewable and Sustainable Energy Reviews*, 25, pp. 470-482

231. Morgan, D.L. (2007) “Paradigms lost and pragmatism regained: methodological implications of combining qualitative and quantitative methods”, *Journal of Mixed Methods Research*, 1 (1), pp. 48-76

232. Muench, S. and Guenther, E. (2013) “A systematic review of bioenergy life cycle assessments”, *Applied Energy*, 112, pp. 257-273

233. Multi-Criteria Decision Making (MCDM) analysis (first) interview with the American Council for an Energy-Efficient Economy (ACEEE), 12/11/2013, Washington DC, USA

234. Multi-Criteria Decision Making (MCDM) analysis (first) interview with the University of Oxford, 11/12/2013, Oxford, UK

235. Multi-Criteria Decision Making (MCDM) analysis (fourth) interview with the University of Oxford, 15/01/2014, Oxford, UK

236. Multi-Criteria Decision Making (MCDM) analysis (second) interview with the American Council for an Energy-Efficient Economy (ACEEE), 12/11/2013, Washington DC, USA

237. Multi-Criteria Decision Making (MCDM) analysis (second) interview with the University of Oxford, 11/12/2013, Oxford, UK

238. Multi-Criteria Decision Making (MCDM) analysis (third) interview with the University of Oxford, 11/12/2013, Oxford, UK

239. Multi-Criteria Decision Making (MCDM) analysis interview with Imperial College London, 19/12/2013, London, UK

240. Multi-Criteria Decision Making (MCDM) analysis interview with Open Energi, 20/01/2014, London, UK

241. Multi-Criteria Decision Making (MCDM) analysis interview with Scottish and Southern Energy (SSE), 14/05/2014, London, UK

242. Multi-Criteria Decision Making (MCDM) analysis interview with the New York State Public Service Commission (NYSPSC), 12/11/2013, Albany, USA

243. Multi-Criteria Decision Making (MCDM) analysis interview with the US Department of Energy (DoE), 12/11/2013, Washington DC, USA

244. Multi-Criteria Decision Making (MCDM) analysis interview with the US Edison Electric Institute (EEI), 12/11/2013, Washington DC, USA

245. Multi-Criteria Decision Making (MCDM) analysis interview with the UK Association for the Conservation of Energy (ACE), 07/01/2014, London, UK

246. Multi-Criteria Decision Making (MCDM) analysis interview with the University of Reading, 09/01/2014, Reading, UK

247. Multi-Criteria Decision Making (MCDM) analysis interview with the UK Office of Gas and Electricity Markets (Ofgem), 16/01/2014, London, UK
248. Multi-Criteria Decision Making (MCDM) analysis interview with the UK Department of Energy and Climate Change (DECC), 17/01/2014, London, UK
249. Multi-Criteria Decision Making (MCDM) analysis interview with the UK Demand Response Association (UK DRA), 06/06/2014, London, UK
250. Mundaca, T.L. and Neij, L. (2010) "A meta-analysis of bottom-up ex-ante energy policy evaluation studies", *In proceedings of the 2010 International Energy Program Evaluation Conference, 'Counting on Energy Programs - It's Why Evaluation Matters',* Paris, France, International Energy Program Evaluation, 2010
251. Murray, M. (2010) "Evaluation of the effectiveness and impact of energy efficiency advertising campaigns", *International Energy Program Evaluation Conference 2010*
252. Nadel, S. and Geller, H. (1996) "Utility DSM: what have we learned? Where are we going?", *Energy Policy*, 24 (4), pp. 289-302
253. Nadel, S. and Gold, R. (2010) "Utility DSM: Off the Coasts and into the Heartland", *The Electricity Journal*, 23 (8), pp. 51-62
254. Nadel, S. and Kushler, M. (2000) "Public Benefit Funds: A Key Strategy for Advancing Energy Efficiency", *The Electricity Journal*, 13 (8), pp. 74-84
255. Nadel, S., Amann, J., Hayes, S., Bin, S., Young, R., Mackres, E., Misuriello, H. and Watson, S. (2013) "An Introduction to US Policies to Improve Building Efficiency, *American Council for an Energy-Efficient Economy (ACEEE)*, Research Report A134, July 2013
256. Nagel, S.S. (2002) *Handbook of Public Policy Evaluation*, Sage Publications, USA
257. National Appliance Energy Conservation Act of 1988, Washington DC: National Congress, USA
258. National Energy Conservation Policy Act of 1978, Washington DC: National Congress, USA
259. National Grid (2015) "Interconnectors", website information can be downloaded from: http://www2.nationalgrid.com/About-us/European-business-development/Interconnectors/ [accessed on 31/03/15]
260. Natural Resources Defense Council (NRDC) (2013) "Decoupling Across the United States", Map, August 2013, can be downloaded from: http://www.nrdc.org/energy/decoupling [accessed on 13/01/14]
261. Neubauer, M., Foster, B., Elliott, N., White, D. and Hornby, R. (2013) "Ohio's Energy Efficient Resource Standard: Impacts on the Ohio Wholesale Electricity Market and Benefits to the State", *American Council for an Energy-Efficient Economy (ACEEE)*, Research Report E138, April 2013
262. Neumann, R. (2013) "Trials and tribulations of joint electric and gas programs - guidelines in integrating joint DSM", *International Energy Program Evaluation Conference 2013*
263. Newborough, M. and Augood, P. (1999) "Demand-side management opportunities for the UK domestic sector", *IEE Proceedings–C Generation, Transmission and Distribution*, 146 (3), pp. 283-293

264. Nicolson, M. (2014) *Can behavioural economics be used to boost registrations to time-of-use electricity tariffs?*, MRes dissertation, University College London (UCL), UK
265. Nishio, K. and Ofuji, K. (2012) "Behavior change and driving forces to save electricity in the electricity crisis in Japan", *International Energy Program Evaluation Conference 2012*
266. O'Connell, N., Pinson, P., Madsen, H. and O'Malley, M. (2014) “Benefits and challenges of electrical demand response – A critical review”, *Renewable and Sustainable Energy Reviews*, 39, pp. 686-699
267. O'Donoghue, T. and Punch K. (2003) *Qualitative Educational Research in Action: Doing and Reflecting*, Routledge, pp. 78
268. O'Drain, M.J. and Edwards, C.M. (2010) "Evaluating low income energy efficiency programs: getting results in California", *International Energy Program Evaluation Conference 2010*
269. Owen, G., Pooley, M. and Ward, J. (2012) “What demand side services could household customers offer?”, *Paper 3 of the GB Electricity Demand – realising the resource project*, Sustainability First, April 2012, UK
270. Owens, S. and Driffil, L. (2008) “How to change attitudes and behaviours in the context of energy”, *Energy Policy*, 36, pp. 4412-4418
271. Pachauri, S., Ürge-Vorsatz, D. and LaBelle, M. (2012) “Synergies between energy efficiency and energy access policies and strategies”, *Global Policy*, 3 (2), pp. 187-197
272. Papathanasopoulou, E., Queirós, A.M., Beaumont, N., Hooper, T. and Nunes, J. (forthcoming) “What are the local impacts of energy systems on marine ecosyetm services”, Collaboration for Environmental Evidence (CEE)
273. Patterson, M.G. (1996) “What is energy efficiency – concepts, indicators and methodological issues”, *Energy Policy*, 24 (5), pp. 377-390
274. Pawson, R. (2002a) “Evidence-based policy – in search of a method”, *Evaluation*, 8 (2), pp. 157-181
275. Pawson, R. (2002b) “Evidence-based policy: the promise of ‘Realist Synthesis’”, *Evaluation*, 8 (3), pp. 340-35
276. Pawson, R. (2006) *Evidence-based Policy: A Realist Perspective*, London: SAGE, UK
277. Pawson, R., Boaz, A., Grayson, L., Long, A. and Barnes, C. (2003) “SCIE Knowledge review 03: Types and quality of knowledge in social care”, *Social Care Institute for Excellence (SCIE)*, November 2003
278. Pengcheng, L., Meng, L., Haihong, C. and Yan, L. (2012) "China's practices on evaluating the energy savings of mandatory energy efficiency standards", *International Energy Program Evaluation Conference 2012*
279. Peplow, M. (2014) “Social sciences suffer from severe publication bias”, *Nature*, News, 28th August 2014
280. Petticrew, M. and Roberts, H. (2006) *Systematic Reviews in the Social Sciences*, Blackwell Publishing, Oxford, UK
281. Poortinga, W. and Pidgeon, N.F. (2003) “Exploring the dimensionality of trust in risk regulation”, *Risk Analysis*, 23 (5), pp. 961-972
282. Porter, J.J., Dessai, S. and Tompkins, E.L. (2014) “What do we know about UK household adaptation to climate change? A systematic review”, *Climate Change*, 127 (2), pp. 371-379

283. Prema, V., Uma Rao, K. and Closepet, A.S. (2014) "A novel predictive DSM strategy to match power outage for optimal cost with solar and diesel power", *Innovative Smart Grid Technologies – Asia (ISGT Asia) Conference*, Institute of Electrical and Electronics Engineers (IEEE), 20-23rd May 2014, Kuala Lumpar, pp. 377-382

284. Prindle, W., Dietsch, N., Elliott, R.N., Kushler, M., Langer, T. and Nadel, S. (2003) "Energy Efficiency's Next Generation: Innovation at the State Level", *American Council for an Energy-Efficient Economy (ACEEE)*, Research Report E031, November 2003

285. Programa Nacional de Conservacão de Energia Eléctrica (PROCEL) of 1985, Brazil

286. Prüggler, W. Prüggler and F. Wirl (2011) "Storage and Demand Side Management as power generator's strategic instruments to influence demand and prices", Energy, 36 (11), pp. 6308-6317

287. Public Utility Regulatory Policy Act of 1978, Washington DC: National Congress, USA

288. Regulatory Assistance Project (RAP) (2012) "Policies to achieve greater energy efficiency", *Global Power Best Practice Series*, October 2012

289. Regulatory Assistance Project (RAP) (2011) *Electricity Regulation in the US: a Guide*, March 2011

290. Rehfuess, E.A., Puzzolo, E., Stanistreet, D., Pope, D. and Bruce, N.G. (2014) "Enablers and barriers to large-scale uptake of improved solid fuel stoves: a systematic review", *Environmental Health Perspectives*, 122 (2), pp. 120-130

291. Renewable Heat Incentive, 2011, Department of Energy and Climate Change (DECC), UK

292. Rogers *et al*. (2000) "Program Theory Evaluation: Practice, Promise, and Problems", *New Directions in Evaluation*, 87, Fall 2000

293. Rogers, E. (1995) *Diffusion of innovations*, 4th edition, New York: Free Press

294. Rose, R. (1991) "What is lesson-drawing", *Journal of Public Policy*, 11, pp. 3-30

295. Rose, R. (1993) *Lesson drawing in public policy: a guide to learning across time and space*, Chatham House

296. Rosenow, J. (2011) "Different paths of change: Home energy efficiency policy in Britain and Germany", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)

297. Rosenow, J. and Galvin, R. (2013) "Evaluating the evaluations: evidence from energy efficiency programmes in Germany and the UK", *Energy and Buildings*, 62, pp. 450-458

298. Rossi, P., Lisey, M. and Freeman, H. (2004) *Evaluation: a systematic approach*, 7th edition, Thousand Oaks: SAGE

299. Sadovnik, R. (2007) "Interpretation and intention in policy analysis", chapter 29 in: Fischer, F., Miller, G.J. and Sidney, M.S. (2007) *Handbook of Public Policy analysis: Theory, Politics, and Methods*, Taylor & Francis Group

300. Saumure, K. and Given, L.M. (2008) "Data Saturation", *The SAGE Encyclopedia of Qualitative Research Methods*, SAGE Publications

301. Saunders, M., Lewis, P. and Thornhill, A. (2007) *Research Methods for Business Students*, 4th Edition, Harlow: Pearson Education Limited

302. Saunders, M., Lewis, P. and Thornhill, A. (2009) *Research Methods for Business Students*, 5th Edition, Harlow: Pearson Education Limited

303. Schultz, D. and Eto, J. (1990) "Carrots and sticks: Shared-savings incentive programs for energy efficiency", *The Electricity Journal*, 3 (10), p. 32-46

304. Schwatz, S. and Carpenter, K.M. (1999) “The right answer for the wrong question: consequences of type III error for public health research”, *American Journal of Public Health*, 89 (8), pp. 1175-1180

305. Sciortino, M., Nowak, S., Witte, P., York, D. and Kushler, M. (2011) "Energy Efficiency Resource Standards: A Progress Report on State Experience", *American Council for an Energy-Efficient Economy (ACEEE)*, Research Report U112, June 2011

306. Shepherd, J., Kavanagh, J., Picot, J., Cooper, K., Harden, A., Barnett-Page, E., Jones, J., Clegg, A., Hartwell, D., Frampton, G.K. and Price, A. (2010) “The effectiveness and cost-effectiveness of behavioural interventions for the prevention of sexually transmitted infections in young people aged 13–19: a systematic review and economic evaluation”, *Health Technology Assessment*, 14 (7)

307. Sioshansi, F. and Vojdani, A. What could possibly be better than real-time pricing? Demand Response, *The Electricity Journal*, June, 2001

308. Smith, S. and Thorne, J. (2003) "An evaluation of the EnergyGuide Label: what we learned", *International Energy Program Evaluation Conference 2003*

309. Smith, V., Devane, D., Begley, C.M. and Clarke, M. (2011) “Methodology in conducting a systematic review of systematic reviews of healthcare interventions”, *BMC Medical Research Methodology*, 11 (15)

310. Snilstveit, B., Oliver, S. and Vojtkova, M. (2012) “Narrative approaches to systematic review and synthesis of evidence for international development policy and practice”, *Journal of Development Effectiveness*, 4 (3), pp. 409-429

311. Sorrell, S. (2007a) “The Rebound Effect: an assessment of the evidence for economy-wide energy savings from improved energy efficiency”, report produced for the UK Energy Research Centre (UKERC), October 2007

312. Sorrell, S. (2007b) “Improving the evidence base for energy policy: the role of systematic reviews”, *Energy Policy*, 35, pp. 1858-1871

313. Sorrell, S. (2014) “Energy substitution, technical change and rebound effects”, *Energies*, 7, pp. 2850-2873

314. Sorrell, S. (2015) “Reducing energy demand: a review of issues, challenges and approaches”, *Renewable and Sustainable Energy Reviews*, 47, pp. 74-82

315. Sorrell, S., Dimitropoulos, J. and Sommerville, M. (2009) “Empirical estimates of the direct rebound effect: a review”, *Energy Policy*, 37, pp. 1356-1371

316. Southern California Edison (SCE) (2007) "2007 Energy Efficiency Annual Report", November 2007, USA

317. Spees, K. and Lave, L.B. (2007) “Demand response and electricity market efficiency”, *The Electricity Journal*, 20 (3), pp. 69-85

318. Staniaszek, D. and Lees, E. (2012) *Determining energy savings for energy efficiency obligation schemes*, European Council for an Energy Efficient Economy (ECEEE) and RAP report, April 2012

319. Steinberger, J.K., van Niel, J. and Bourg, D. (2009) “Profiting from negawatts: reducing absolute consumption and emissions through a performance-based energy economy”, *Energy Policy*, 37, pp. 361-370

320. Stern, F. and Vantzis, D. (2014) “Protocols for Evaluating Energy Efficiency – Both Sides of the Atlantic”, *International Energy Policy and Programme Evaluation Conference*, 9-11th September 2014, Berlin, Germany

321. Strachan, N. and Warren, P. (2011) “Incorporating behavioural complexity in energy-economic models”, UK Energy Research Centre Conference on: *Energy and People: Futures, Complexity and Challenges*, 20-21 September 2011, Environmental Change Institute, Oxford, UK

322. Strbac, G. (2008) “Demand side management: benefits and challenges”, *Energy Policy*, 36, pp. 4419-4426

323. Stromback, J., Dromacque, C. and Yassin, M.H. *The potential of smart meter enabled programs to increase energy and systems efficiency: a mass pilot comparison*, Vaasaett, 2011, pp. 13

324. Sumi, D. and Prahl, R. (2001) "A comprehensive examination of the market effects of a public benefits-sponsored pilot program: lessons learned from Wisconsin's Focus on Energy", *International Energy Program Evaluation Conference 2001*

325. Suna, D. and Haas, R. (2013) "How to calculate energy savings and costs of energy saving obligations in a harmonized way?", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)

326. Taylor, B., Trombley, D. and Reinaud, J. (2012) "Energy Efficiency Resource Acquisition Program Models in North America", *American Council for an Energy-Efficient Economy (ACEEE)*, Research Report IE126, November 2012

327. Thaler, R.H. and Sunstein, C.R. (2008) *Nudge: Improving Decisions About Health, Wealth, and Happiness*, Yale University Press

328. THINK (2012) “How to refurbish all buildings by 2050”, Topic 7, Final report, THINK, June 2012

329. Tholen, L., Kiyar, D., Venjakob, M., Xia, C., Thomas, S. and Aydin, V. (2013) "What makes a good policy? Guidance for assessing and implementing energy efficiency policies", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)

330. Thomas, S., Adnot, J., Alari, P., Irrek, W., Lopes, C., Nilsson, L.J., Pagliano, L., Verbruggen, A. (2000) “Completing the market for least-cost energy services”, Report, prepared by Wuppertal Institute for Climate Environment Energy *et al*.

331. Thornley, P., Upham, P. and Tomei, J. (2009) “Sustainability constraints on UK bioenergy development”, *Energy Policy*, 37 (12), pp. 5623-5635

332. Titus, E., Michals, J., Hurley, D., Osann, E.R. and Waite, S. (2009) "Energy efficiency as a resource in the PJM Capacity Market", *International Energy Program Evaluation Conference 2009*

333. Togeby, M., Dyhr-Mikkelsen, K., Larsen, A., Hansen, M.J. and Bach, P. (2009) "Danish energy efficiency policy: revisited and future improvements", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)

334. Torriti, J., Hassan, M.G. and Leach, M. (2010) “Demand response experience in Europe: policies, programmes and implementation”, *Energy*, 35 (4), pp. 1575-1583

335. UK Civil Service (2014) “What is a rapid evidence assessment?”, UK Civil Service website, accessed on 27/11/14, can be downloaded from: http://www.civilservice.gov.uk/networks/gsr/resources-and-guidance/rapid-evidence-assessment/what-is

336. UK Department of Communities and Local Government (CLG) (2009) *Multi-criteria analysis: a manual*, January 2009

337. UK Department of Energy and Climate Change (DECC) (2010) *The Green Deal: a summary of the Government’s proposals*, Government policy document, Department of Energy and Climate Change (DECC)

338. UK Department of Energy and Climate Change (DECC) (2011a) "Evaluation of the Low Carbon Buildings Programme", Research Report, August 2011, DECC

339. UK Department of Energy and Climate Change (DECC) (2011b) "Evaluation of the Community Energy Saving Programme", Research Report, October 2011, DECC

340. UK Department of Energy and Climate Change (DECC) (2011c) "Energy supplier obligation policies: evaluation synthesis", Research Report, October 2011, DECC

341. UK Department of Energy and Climate Change (DECC) (2014a) “D3: Opportunities for integrating demand side energy policies”, Government research report, July 2014

342. UK Department of Energy and Climate Change (DECC) (2014b) *Smart metering implementation programme: smart metering equipment technical specifications*, Version 1.58, Draft document, 28th November 2014

343. UK Department of Energy and Climate Change (DECC) (2014c) “Electricity demand reduction pilot”, Guidance documents can be downloaded from: https://www.gov.uk/electricity-demand-reduction-pilot [accessed on 18/03/15]

344. UK Department of Energy and Climate Change (DECC) (2014d) *Smart metering implementation programme: third annual report on the roll-out of smart meters*, December 2014

345. UK Office of Gas and Electricity Markets (Ofgem) (2014) *State of the Market Assessment*, Assessment Report, 27/03/14

346. UK Office of Gas and Electricity Markets (Ofgem) (2010) *Smart Metering Implementation Programme: Statement of Design Requirements*, Supporting Document for Consultation Response, 27/07/10

347. UK Parliament (2011) *Briefing on Interconnectors*, London

348. University of Illinois, *Levels of evidence*, can be downloaded from: http://ebp.lib.uic.edu/nursing/node/12 [accessed on 07/11/13]

349. Ürge-Vorsatz, D. and Hauff, J. (2001) “Drivers of market transformation: analysis of the Hungarian lighting success story”, *Energy Policy*, 29, pp. 801-810

350. Ürge-Vorsatz, D., Cabeza, L.F., Serrano, S., Barreneche, C. and Petrichenko, K. (2015) “Heating and cooling energy trends and drivers in buildings”, *Renewable and Sustainable Energy Reviews*, 41, pp. 85-98
351. Ürge-Vorsatz, D., Koeppel, S. and Mirasgedis, S. (2007) “Appraisal of policy instruments for reducing buildings’ CO_2 emissions”, *Building Research & Information*, 35 (4), pp. 458-477
352. Utilities Act of 2000, London: HMSO, UK
353. Vandenbroucke, J.P. (2001) “In defense of case reports and case series”, *Annals of Internal Medicine*, 134 (4), pp. 330-334
354. Verbruggen, A. (2008) “Renewable and nuclear power – a common future?”, *Energy Policy*, 36 (11), pp. 4036-4047
355. Vermont Public Service Department (2014) “Energy Efficiency”, can be downloaded from: http://publicservice.vermont.gov/topics/energy_efficiency [accessed on 07/11/14]
356. Vine, E., Rhee, C.H. and Lee, K.D. (2006) "Measurement and evaluation of energy efficiency programs: California and South Korea", *Energy*, 31 (6-7), pp. 1100-1113
357. Vreuls, H. (2014) “Impact evaluation of energy efficiency and DSM programmes”, webinar, IEA DSM University, Leonardo Energy
358. Waide, P. and Buchner, B. (2008) "Utility energy efficiency schemes: savings obligations and trading", *Energy Efficiency*, 1 (4), pp. 297-311
359. Walawalkar, R., Fernands, S., Thakur, N. and Chevva, K.R. (2010) "Evolution and current status of demand response (DR) in electricity markets: Insights from PJM and NYISO", *Energy*, 35 (4), pp. 1553-1560
360. Walsh, P.R. and Todeva, E. (*year not stated*) “Vertical and horizontal integration in the utilities sector: the case of RWE”, *unpublished paper*, University of Surrey, UK
361. Wang, J., Bloyd, C.N., Hu, Z. and Tan, Z. (2010) "Demand response in China", *Energy*, 35 (4), pp. 1592-1597
362. Warm Front, 2001, Department of Trade and Industry (DTI), UK
363. Warm Homes and Energy Conservation Act of 2000, London: HMSO, UK
364. Warren, P. (2014a) “A review of demand-side management policy in the UK”, *Renewable and Sustainable Energy Reviews*, 29, pp. 941-951
365. Warren, P. (2014b) “The use of systematic reviews to analyse demand-side management policy”, *Energy Efficiency*, 7, pp. 417-427
366. Warren, P. (under review) “Political Economy of Demand-Side Management”, *Economics of Energy and Environmental Policy*, International Association for Energy Economics (IAEE)
367. Watson, J., Byrne, R., Morgan Jones, M., Tsang, F., Opazo, J., Fry, C. and Castle-Clarke, S. (2012) “What are the major barriers to increased use of modern energy services among the world’s poorest people, and are interventions to overcome these effective?” Project Report, Collaboration for Environmental Evidence (CEE)
368. White Paper (1989) This common inheritance, Environment White Paper, London: HMSO, UK
369. White Paper (2007) Meeting the energy challenge, Energy White Paper, London: HMSO, UK

370. Whitmarsh, L. (2011) "Social and psychological drivers of energy consumption behaviour and energy transitions", Chapter 11, in: Dietz, S., Michie, J. and Oughton, C. (2011) *The Political Economy of the Environment: an Interdisciplinary Approach*, pp. 213-228, Routledge: Oxon
371. Wikler, G.A. (2000) "Policy Options for Energy Efficiency Initiatives", *The Electricity Journal*, 13 (1), pp. 61-68
372. Wilson, C. and Dowlatabadi, H. (2007) "Models of Decision Making and Residential Energy Use", *Annual Review of Environment and Resources*, 32, pp. 169-203
373. Wirtshafter, R.M., Harper, B., Faesy, R., Reed, G., Badger, L., Chiodo, J., Killian, E. and George, K. (2011) "The costs and benefits of measuring if States meet 90% compliance with Building Codes", *International Energy Program Evaluation Conference 2011*
374. Wirtshafter, R.M., Parlin, K., Hungerford, D., McKinley, K. and Bordner, R. (2007) "Desperately seeking savings from small scale demand response: the California experience", *International Energy Program Evaluation Conference 2007*
375. Wittneben, B.B.F. (2012) "The impact of the Fukushima nuclear accident on European energy policy", *Environmental Science & Policy*, 15, pp. 1-3
376. World Bank (2011) *Energy use (kg of oil equivalent per capita)*, accessed on 09/10/14, can be downloaded from: http://data.worldbank.org/indicator/EG.USE.PCAP.KG.OE
377. World Bank (2014) *Climate Change Knowledge Portal*, 1990-2009 averaged data from the Climate Research Unit, University of East Anglia
378. Wüstenhagen, R., Wolsink, M. and Bürer, M.J. (2007) "Social acceptance of renewable energy innovation: an introduction to the concept", *Energy Policy*, 35 (5), pp. 2683-2691
379. Yacobucci, B.D. (2014) "Energy Policy: 113th Congress Issues", *Congressional Research Service*, 18th September 2014
380. Yang, K. (2007) "Quantitative methods for policy analysis", chapter 23 in: Fischer, F., Miller, G.J. and Sidney, M.S. (2007) *Handbook of Public Policy analysis: Theory, Politics, and Methods*, Taylor & Francis Group, USA
381. Yang, M. and Rumsey, P. (1997) "Energy conservation in typical Asian countries", *Energy Sources*, 19 (5), pp. 507-521
382. Yanow, D. (2007) "Qualitative-interpretive methods in policy research", chapter 27 in: Fischer, F., Miller, G.J. and Sidney, M.S. (2007) *Handbook of Public Policy analysis: Theory, Politics, and Methods*, Taylor & Francis Group, USA
383. Yu, Y. (2010) "Policy redesign for solving the financial bottleneck in demand side management (DSM) in China", *Energy Policy*, 38 (10), pp. 6101-6110
384. Yu, Y. (2012) "How to fit demand side management (DSM) into current Chinese electricity system reform?", *Energy Economics*, 34 (2), pp. 549-557
385. Yun, J. and Price, L. (2011) "Voluntary energy efficiency agreements in China: history, impact, and future", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)

386. Zarnikau, J.W. (2010) "Demand participation in the restructured Electric Reliability Council of Texas market", *Energy*, 35 (4), pp. 1536-1543
387. Zheng, N., Zhou, N., Fino-Chen, C. and Fridley, D. (2012) "Evaluation of local enforcement of energy efficiency standards and labeling program in China", *International Energy Program Evaluation Conference 2012*
388. Zhou, N., Mcneil, M. and Levine, M. (2011) "Assessment of Building Energy-Saving Policies and Programs in China during the 11th Five Year Plan", Ernest Orlando Lawrence Berkeley National Laboratory, USA
389. Zuckerman, J., Deason, J. and Chandrasehkeran, S. (2013) "Rewarding efficiency: lessons from California's shareholder incentives", *International Energy Program Evaluation Conference 2013*

7.2 Documents Included in the Systematic Review

1. Agnew, K., Burke, R. and Ham-su, P. (2009) "Participation of demand response resources in ISO New England's Ancillary Service Markets", *International Energy Program Evaluation Conference 2009*
2. Arsenault, E., Bernard, J.-T. and Genest-Laplante, E. (1996) "Hydro-Québec energy savings programs: ‘Watt’ are they worth?", *Resource and Energy Economics*, 18 (1), pp. 65-81
3. Association for the Conservation of Energy (ACE) (2013) *Financing energy efficiency in buildings: an international review of best practice and innovation*, A report to the World Energy Council, October 2013
4. Atanasiu, B. and Constantinescu, T. (2011) "A comparative analysis of the energy performance certificates schemes within the European Union: Implementing options and policy recommendations", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)
5. Bachrach, D. (2003) "Energy Efficiency Leadership in California: Preventing the Next Crisis", *The Electricity Journal*, 16 (6), pp. 37-47
6. Bertoldi, P. and Rezessy, S. (2007) "Assessment of White Certificate Schemes and their energy saving evaluation methods", *International Energy Program Evaluation Conference 2007*
7. Bertoldi, P. and Rezessy, S. (2008) "Tradable white certificate schemes: fundamental concepts", *Energy Efficiency*, 1, pp. 237-255
8. Betz, R., Jones, M.C., MacGill, I. and Passey, R. (2013) "Trading in energy efficiency in Australia: What are the lessons learnt so far?', *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)
9. Bin, S. and Jun, L. (2012) "Building Energy Efficiency Policies in China", *American Council for an Energy-Efficient Economy (ACEEE)*, Research Report E129, June 2012
10. Broc, J.-S., Melo, C.A. and Jannuzzi, G. (2012) "Detailed comparison of Brazilian and French obligation schemes to promote energy efficiency", *International Energy Program Evaluation Conference 2012*

11. Brown, M.A. and Mihlmester, P.E. (1995) "What has DSM achieved in California?", report number CONF-950643-1, Department of Energy (DoE), USA
12. Bukarica, V., Debrecin, N., Borkovic, Z.H. and Pesut, D. (2012) "Evaluating energy savings arising from NEEAP implementation: lessons learned in Croatia", *International Energy Program Evaluation Conference 2012*
13. Bundgaard, S.S., Dyhr-Mikkelsen, K., Larsen, A.E. and Togeby, M. (2013) "Energy Efficiency Obligation Schemes in the EU - Lessons Learned from Denmark", First quarter 2013, *International Association for Energy Economics (IAEE)*
14. Bundgaard, S.S., Togeby, M., Dyhr-Mikkelsen, K., Sommer, T., Kajaerbye, V.H. and Larsen, A.E. (2013) "Spending to save: evaluation of the energy efficiency obligation in Denmark", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)
15. Cappers, P., Goldman, C. and Kathan, D. (2010) "Demand response in U.S. electricity markets: Empirical evidence", *Energy*, 35 (4), pp. 1526-1535
16. Cappers, P., Goldman, C.A. and Kathan, D. (2009) "Demand Response in US Electricity Markets: Empirical Evidence", report number LBNL-2124E, Ernest Orlando Lawrence Berkeley National Laboratory, USA
17. Centre for Strategy & Evaluation Services (2012) "Evaluation of the Ecodesign Directive", Final report, UK
18. Colby, J. and Davis, S. (2011) "Can short term ARRA stimulus funding achieve long term market transformation?", *International Energy Program Evaluation Conference 2011*
19. Crossley, D. (2008) "Tradeable energy efficiency certificates in Australia", *Energy Efficiency*, 1, pp. 267-281
20. Crossley, D. (2010) "International best practice in using energy efficiency and demand management to support electricity networks", *Report 4 - Scaling the Peaks: Demand Management and Electricity Networks*, Australian Alliance to Save Energy, Australia
21. Di Santo, D., Forni, D., Venyurini, V. and Biele, E. (2011) "The white certificate scheme: the Italian experience and proposals for improvement", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)
22. Doris, E., Cochran, J. and Vorum, M. (2009) "Energy Efficiency Policy in the United States: Overview of Trends at Different Levels of Government", report number NREL/TP-6A2-46532, National Renewable Energy Laboratory, Department of Energy (DoE), USA
23. Drew, T. (2009) "An assessment of California's energy efficiency incentive mechanism", *International Energy Program Evaluation Conference 2009*
24. Eyre, N. (2008) "Regulation of energy suppliers to save energy - lessons from the UK debate", *British Institute of Energy Economists (BIEE)*
25. Eyre, N., Pavan, M. and Bodineau, L. (2009) "Energy company obligations to save energy in Italy, the UK and France: what have we learnt?", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)

26. Fleiter, T., Gruber, E., Eichhammer, W. and Worrell, E. (2012) "The German energy audit program for firms – a cost-effective way to improve energy efficiency?", *Energy Efficiency*, 5, pp. 447-469
27. Friedmann, R. and Sheinbaum, C. (1998) "Mexican electric end-use efficiency: experiences to date", *Annual Review of Energy and the Environment*, 23, pp. 225-252
28. Gillich, A. and Sunikka-Blank, M. (2013) "Barriers to domestic energy efficiency – an evaluation of retrofit policies and market transformation strategies", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)
29. Gillingham, K., Newell, R. and Palmer, K. (2006) "Energy efficiency policies: a retrospective examination", *Annual Review of Environment and Resources*, 31, pp. 161-192
30. Giraudet, L.-G., Bodineau, L. and Finon, D. (2012) "The costs and benefits of white certificates schemes", *Energy Efficiency*, 5, pp. 179-199
31. Gold, R. and Nadel, S. (2011) "Assessing the Harvest: Implementation of the Energy Efficiency Provisions in the Energy Policy Act of 2005", *American Council for an Energy-Efficient Economy (ACEEE)*, Research Report E113, March 2011
32. Gold, R. and Nadel, S. (2011) "Energy Efficiency Tax Incentives, 2005-2011: How Have They Performed?", *American Council for an Energy-Efficient Economy (ACEEE)*, White Paper on Tax Incentives, June 2011
33. Gold, R., Nadel, S., Laitner, J.A. and deLaski, A. (2011) "Appliance and Equipment Efficiency Standards: A Moneymaker and Job Creator", *American Council for an Energy-Efficient Economy (ACEEE)*, Research Report A111, January 2011
34. Goldman, C.A. and Kito, M.S. (1994) "Review of Demand-Side Bidding Programs: Impacts, Costs and Cost-Effectiveness", report number LBL-35021, Lawrence Berkeley National Laboratory and the University of California, USA
35. Goldman, C.A., Stuart, E., Hoffman, I., Fuller, M.C. and Billingsley, M.A. (2011) "Interactions between Energy Efficiency Programs funded under the Recovery Act and Utility Customer-Funded Energy Efficiency Programs", report number LBNL-4322E, Ernest Orlando Lawrence Berkeley National Laboratory, USA
36. Gruber, E., Fleiter, T., Mai, M. and Frahm, B.-J. (2011) "Efficiency of an energy audit programme for SMEs in Germany – results of an evaluation study", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)
37. Guertler, P., Royston, S. and Robson, D. (2013) "Somewhere between a ‘Comedy of errors’ and ‘As you like it’? A brief history of Britain’s ‘Green Deal’ so far", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)
38. Hadley, S. and Hirst, E. (1995) "Utility DSM programs from 1989 through 1998: continuation or cross roads?", report number ORNL/CON-405, Oak Ridge National Laboratory, USA
39. Harmelink, M., Nilsson, L. and Harmsen, R. (2008) "Theory-based policy evaluation of 20 energy efficiency instruments", *Energy Efficiency*, 1, pp. 131-148
40. Heffner, G.C., Lees, E., Crossley, D. and Migden-Ostrander, J. (2012) "Evaluating policies for energy provider-delivered energy efficiency", *International Energy Program Evaluation Conference 2012*

41. Hong, T. (2009) "A close look at the China Design Standard for Energy Efficiency of Public Buildings", *Energy and Buildings*, 41, pp. 426-435
42. Horowitz (2013) "Purchased energy and policy impacts in the US manufacturing sector", *Energy Efficiency*, April 2013
43. IEA DSM Programme (2004) "INDEEP Analysis Report 2004", *Task 1 Subtask 8 - International Database on Demand-Side Management Technologies and Programmes*, International Energy Agency (IEA)
44. IEA DSM Programme (2005) "Evaluation Guidebook Volume 2 - Country Reports and Case Examples Used for the Evaluation Guidebook", *Task 1 Subtask 9 - Evaluation Guidebook on the impact of DSM and Energy Efficiency Programmes for Kyoto's GHG Targets*, International Energy Agency (IEA)
45. IEA DSM Programme (2006) "Task 14 Final Report", *Task 14 - Market Mechanisms for White Certificate Trading*, International Energy Agency (IEA)
46. IEA DSM Programme (2008) "Evaluation and Acquisition of Network-driven DSM Resources", *Task 15 - Network driven DSM*, International Energy Agency (IEA)
47. International Institute for Energy Conservation (IIEC) *Demand Side Management: Best Practices Guidebook for Pacific Island Power Utilities*, Report prepared for the South Pacific Applied Geoscience Commission (SOPAC) and the United Nations Department of Economic and Social Affairs (UNDESA)
48. Kimura, O. (2009) "Is public R&D in energy efficiency really effective? – a case in Japan and its implications", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)
49. Kushler, M. and Vine, E. (2003) "Examining California's Energy Efficiency Policy Response to the 2000/2001 Electricity Crisis: Practical Lessons Learned Regarding Policies, Administration, and Implementation", *American Council for an Energy-Efficient Economy (ACEEE)*, Research Report U033, March 2003
50. Kushler, M., Nowak, S. and Witte, P. (2012) "A National Survey of State Policies and Practices for the Evaluation of Ratepayer-Funded Energy Efficiency Programs", *American Council for an Energy-Efficient Economy (ACEEE)*, Research Report U122, February 2012
51. Kushler, M., York, D. and Witte, P. (2004) "Five Years In: An Examination of the First Half-Decade of Public Benefits Energy Efficiency Policies", *American Council for an Energy-Efficient Economy (ACEEE)*, Research Report U041, April 2004
52. Lee, A.D., Onisko, S.A., Sandahl, L.J. and Butler, J. (1994) "Everyone wins! — A program to upgrade energy efficiency in manufactured housing", *The Electricity Journal*, 7 (2), pp. 77-87
53. Lees, E. (2006) "Evaluation of the Energy Efficiency Commitment 2002-2005", Report to DEFRA, prepared by Eoin Lees Energy, 28th February 2006, UK
54. Lees, E. (2008) "Evaluation of the Energy Efficiency Commitment 2005-2008", Report to DEFRA, prepared by Eoin Lees Energy, 28th February 2006, UK
55. Lepetitgaland, K.K., Vasconcellos, L.E.M., Magalhaes, L.P., Costa, A.L.P., Costa, L.F.T. and Friedmann, R. (2011) "The strategic value of

evaluation for Brazil and Neighbouring countries: the experience of Procel", *International Energy Program Evaluation Conference 2011*

56. Lowry, M., Irwin, S. and Waeckerlin, E. (2004) "Demand Response in the West: Lessons for States and Provinces", report prepared for the Department of Energy (DoE), USA
57. Mallaburn, P.S. and Eyre, N. (2013) "Lessons from energy efficiency policy and programmes in the UK from 1973 to 2013", *Energy Efficiency*, February 2013
58. McRae, M., James, A., Kim, A. and Spahic, M. (2011) "Measuring diffusion in a market transformation program", *International Energy Program Evaluation Conference 2011*
59. Meyers, E.M. and Hu, G.M. (1999) "Demand-Side Carbon Reduction Strategies in an Era of Electric Industry Competition", *The Electricity Journal*, 12 (1), pp. 72-81
60. Multi-Criteria Decision Making (MCDM) analysis (first) interview with the American Council for an Energy-Efficient Economy (ACEEE), 12/11/2013, Washington DC, USA
61. Multi-Criteria Decision Making (MCDM) analysis (first) interview with the University of Oxford, 11/12/2013, Oxford, UK
62. Multi-Criteria Decision Making (MCDM) analysis (fourth) interview with the University of Oxford, 15/01/2014, Oxford, UK
63. Multi-Criteria Decision Making (MCDM) analysis (second) interview with the American Council for an Energy-Efficient Economy (ACEEE), 12/11/2013, Washington DC, USA
64. Multi-Criteria Decision Making (MCDM) analysis (second) interview with the University of Oxford, 11/12/2013, Oxford, UK
65. Multi-Criteria Decision Making (MCDM) analysis (third) interview with the University of Oxford, 11/12/2013, Oxford, UK
66. Multi-Criteria Decision Making (MCDM) analysis interview with Imperial College London, 19/12/2013, London, UK
67. Multi-Criteria Decision Making (MCDM) analysis interview with Open Energi, 20/01/2014, London, UK
68. Multi-Criteria Decision Making (MCDM) analysis interview with Scottish and Southern Energy (SSE), 14/05/2014, London, UK
69. Multi-Criteria Decision Making (MCDM) analysis interview with the New York State Public Service Commission (NYSPSC), 12/11/2013, Albany, USA
70. Multi-Criteria Decision Making (MCDM) analysis interview with the US Department of Energy (DoE), 12/11/2013, Washington DC, USA
71. Multi-Criteria Decision Making (MCDM) analysis interview with the US Edison Electric Institute (EEI), 12/11/2013, Washington DC, USA
72. Multi-Criteria Decision Making (MCDM) analysis interview with the UK Association for the Conservation of Energy (ACE), 07/01/2014, London, UK
73. Multi-Criteria Decision Making (MCDM) analysis interview with the University of Reading, 09/01/2014, Reading, UK
74. Multi-Criteria Decision Making (MCDM) analysis interview with the UK Office of Gas and Electricity Markets (Ofgem), 16/01/2014, London, UK
75. Multi-Criteria Decision Making (MCDM) analysis interview with the UK Department of Energy and Climate Change (DECC), 17/01/2014, London, UK

76. Multi-Criteria Decision Making (MCDM) analysis interview with the UK Demand Response Association (UK DRA), 06/06/2014, London, UK
77. Murray, M. (2010) "Evaluation of the effectiveness and impact of energy efficiency advertising campaigns", *International Energy Program Evaluation Conference 2010*
78. Nadal, S., Amann, J., Hayes, S., Bin, S., Young, R., Mackres, E., Misuriello, H. and Watson, S. (2013) "An Introduction to US Policies to Improve Building Efficiency, *American Council for an Energy-Efficient Economy (ACEEE)*, Research Report A134, July 2013
79. Nadel, S. and Gold, R. (2010) "Utility DSM: Off the Coasts and into the Heartland", *The Electricity Journal*, 23 (8), pp. 51-62
80. Nadel, S. and Kushler, M. (2000) "Public Benefit Funds: A Key Strategy for Advancing Energy Efficiency", *The Electricity Journal*, 13 (8), pp. 74-84
81. Neubauer, M., Foster, B., Elliott, N., White, D. and Hornby, R. (2013) "Ohio's Energy Efficient Resource Standard: Impacts on the Ohio Wholesale Electricity Market and Benefits to the State", *American Council for an Energy-Efficient Economy (ACEEE)*, Research Report E138, April 2013
82. Neumann, R. (2013) "Trials and tribulations of joint electric and gas programs - guidelines in integrating joint DSM", *International Energy Program Evaluation Conference 2013*
83. Nishio, K. and Ofuji, K. (2012) "Behavior change and driving forces to save electricity in the electricity crisis in Japan", *International Energy Program Evaluation Conference 2012*
84. O'Drain, M.J. and Edwards, C.M. (2010) "Evaluating low income energy efficiency programs: getting results in California", *International Energy Program Evaluation Conference 2010*
85. Pengcheng, L., Meng, L., Haihong, C. and Yan, L. (2012) "China's practices on evaluating the energy savings of mandatory energy efficiency standards", *International Energy Program Evaluation Conference 2012*
86. Prindle, W., Dietsch, N., Elliott, R.N., Kushler, M., Langer, T. and Nadel, S. (2003) "Energy Efficiency's Next Generation: Innovation at the State Level", *American Council for an Energy-Efficient Economy (ACEEE)*, Research Report E031, November 2003
87. Regulatory Assistance Project (RAP) (2012) "Policies to achieve greater energy efficiency", *Global Power Best Practice Series*, October 2012
88. Rosenow, J. (2011) "Different paths of change: Home energy efficiency policy in Britain and Germany", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)
89. Rosenow, J. and Galvin, R. (2013) "Evaluating the evaluations: evidence from energy efficiency programmes in Germany and the UK", *Energy and Buildings*, 62, pp. 450-458
90. Schultz, D. and Eto, J. (1990) "Carrots and sticks: Shared-savings incentive programs for energy efficiency", *The Electricity Journal*, 3 (10), pp. 32
91. Sciortino, M., Nowak, S., Witte, P., York, D. and Kushler, M. (2011) "Energy Efficiency Resource Standards: A Progress Report on State Experience", *American Council for an Energy-Efficient Economy (ACEEE)*, Research Report U112, June 2011

92. Smith, S. and Thorne, J. (2003) "An evaluation of the EnergyGuide Label: what we learned.", *International Energy Program Evaluation Conference 2003*
93. Southern California Edison (SCE) (2007) "2007 Energy Efficiency Annual Report", November 2007, USA
94. Spees, K. and Lave, L.B. (2007) "Demand Response and Electricity Market Efficiency", *The Electricity Journal*, 20 (3), pp. 69-85
95. Sumi, D. and Prahl, R. (2001) "A comprehensive examination of the market effects of a public benefits-sponsored pilot program: lessons learned from Wisconsin's Focus on Energy", *International Energy Program Evaluation Conference 2001*
96. Suna, D. and Haas, R. (2013) "How to calculate energy savings and costs of energy saving obligations in a harmonized way?", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)
97. Taylor, B., Trombley, D. and Reinaud, J. (2012) "Energy Efficiency Resource Acquisition Program Models in North America", *American Council for an Energy-Efficient Economy (ACEEE)*, Research Report IE126, November 2012
98. Tholen, L., Kiyar, D., Venjakob, M., Xia, C., Thomas, S. and Aydin, V. (2013) "What makes a good policy? Guidance for assessing and implementing energy efficiency policies", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)
99. Titus, E., Michals, J., Hurley, D., Osann, E.R. and Waite, S. (2009) "Energy efficiency as a resource in the PJM Capacity Market", *International Energy Program Evaluation Conference 2009*
100. Togeby, M., Dyhr-Mikkelsen, K., Larsen, A., Hansen, M.J. and Bach, P. (2009) "Danish energy efficiency policy: revisited and future improvements", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)
101. UK Department of Energy and Climate Change (DECC) (2011) "Energy supplier obligation policies: evaluation synthesis", Research Report, October 2011, DECC
102. UK Department of Energy and Climate Change (DECC) (2011) "Evaluation of the Community Energy Saving Programme", Research Report, October 2011, DECC
103. UK Department of Energy and Climate Change (DECC) (2011) "Evaluation of the Low Carbon Buildings Programme", Research Report, August 2011, DECC
104. UK Department of Energy and Climate Change (DECC) (2011) "Low Carbon Building Programme 2006-2011", Final Report, August 2011, DECC
105. Vine, E., Rhee, C.H. and Lee, K.D. (2006) "Measurement and evaluation of energy efficiency programs: California and South Korea", *Energy*, 31 (6-7), pp. 1100-1113
106. Waide, P. and Buchner, B. (2008) "Utility energy efficiency schemes: savings obligations and trading", *Energy Efficiency*, 1 (4), pp. 297-311
107. Walawalkar, R., Fernands, S., Thakur, N. and Chevva, K.R. (2010) "Evolution and current status of demand response (DR) in

electricity markets: Insights from PJM and NYISO", *Energy*, 35 (4), pp. 1553-1560
108. Wang, J., Bloyd, C.N., Hu, Z. and Tan, Z. (2010) "Demand response in China", *Energy*, 35 (4), pp. 1592-1597
109. Wikler, G.A. (2000) "Policy Options for Energy Efficiency Initiatives", *The Electricity Journal*, 13 (1), pp. 61-68
110. Wirtshafter, R.M., Harper, B., Faesy, R., Reed, G., Badger, L., Chiodo, J., Killian, E. and George, K. (2011) "The costs and benefits of measuring if States meet 90% compliance with Building Codes", *International Energy Program Evaluation Conference 2011*
111. Wirtshafter, R.M., Parlin, K., Hungerford, D., McKinley, K. and Bordner, R. (2007) "Desperately seeking savings from small scale demand response: the California experience", *International Energy Program Evaluation Conference 2007*
112. Yang, M. and Rumsey, P. (1997) "Energy conservation in typical Asian countries", *Energy Sources*, 19 (5), pp. 507-521
113. Yu, Y. (2010) "Policy redesign for solving the financial bottleneck in demand side management (DSM) in China", *Energy Policy*, 38 (10), pp. 6101-6110
114. Yu, Y. (2012) "How to fit demand side management (DSM) into current Chinese electricity system reform?", *Energy Economics*, 34 (2), pp. 549-557
115. Yun, J. and Price, L. (2011) "Voluntary energy efficiency agreements in China: history, impact, and future", *ECEEE Summer Study 2013 Proceedings*, European Council for an Energy-Efficient Economy (ECEEE)
116. Zarnikau, J.W. (2010) "Demand participation in the restructured Electric Reliability Council of Texas market", *Energy*, 35 (4), pp. 1536-1543
117. Zheng, N., Zhou, N., Fino-Chen, C. and Fridley, D. (2012) "Evaluation of local enforcement of energy efficiency standards and labeling program in China", *International Energy Program Evaluation Conference 2012*
118. Zhou, N., Mcneil, M. and Levine, M. (2011) "Assessment of Building Energy-Saving Policies and Programs in China during the 11th Five Year Plan", Ernest Orlando Lawrence Berkeley National Laboratory, USA
119. Zuckerman, J., Deason, J. and Chandrasehkeran, S. (2013) "Rewarding efficiency: lessons from California's shareholder incentives", *International Energy Program Evaluation Conference 2013*

8 Appendix

UCL ENERGY INSTITUTE

UCL

Instructions for Systematic Review Pilot

Research questions:

1) What DSM policies have been implemented around the world with high quality documented evaluations?
2) How and why do DSM policies succeed or fail, and what policies have been successful?

**These are provided for background purposes only*

Required resources:

- All the data and information can be put into the spreadsheet provided
- Internet connection and access to the electronic database of the journal *Resource and Energy Economics*
- The process should take about a half-day to do and there are six main stages

Process:

1. Input the search term "Demand-side management" AND programmes AND policies into the electronic database search bar of the journal, *Energy*, and record the total number of hits returned in the spreadsheet provided
2. Skim the titles and exclude all those that do not appear relevant (i.e. do not evaluate DSM policies) and record the new number of filtered hits in the spreadsheet – documents should relate to one or more of the following policies:

Demand-Side Management Policies
Incentive payment-based demand response
Price-based demand response
Infrastructure rollouts (e.g. smart meter rollouts)
Utility obligations
Market transformations (e.g. for energy efficiency)
Loans and subsidies (including tax incentives)
Research and development programmes (only large-scale)
Labelling (appliances, equipment, or buildings)
Performance standards (appliances, equipment, or buildings)
Information campaigns (including energy audits)
Utility business models (e.g. decoupling, performance targets)
Voluntary programmes

3. Skim the Abstracts of the filtered list and further reduce the number of documents by relevance – apply the following inclusion and exclusion criteria to do so, and then record the new number of documents in the spreadsheet

Inclusion Criteria
Documents that answer the research questions
Documents that pass the study quality assessment scale
Documents that discuss government-stimulated policies and programmes
Documents that are written in English
Documents that are freely accessible and downloadable from the internet

Exclusion Criteria
Documents that look at DSM policy but not mechanisms and impacts
Documents that do not pass the study quality assessment scale
Documents that discuss utility-stimulated DSM programmes
Documents that discuss trials, pilots, and small-scale R&D programmes
Documents that model the future potential of DSM
Documents that discuss theoretical aspects of DSM policy
Documents that are not written in English
Documents from hand searching
Documents from referrals
Documents from bibliographies and 'snow-balling'

4. To obtain the final sample of documents that are of high study quality, subject the remaining documents to the following study quality assessment scale (as shown in the spreadsheet):

- **4 points:** Has the process for policy implementation been clearly explained?
  - *1 point: Have details on the policy implementer been given?*
  - *1 point: Have details on how the policy was designed been given?*
  - *1 point: Have details on how the policy was implemented been given?*
  - *1 point: Have details on why the policy was implemented been given?*
- **4 points:** Has the process for policy evaluation been clearly explained?
  - *1 point: Have details on the policy evaluator been given?*
  - *1 point: Have details on how the policy was evaluated been given?*
  - *1 point: Have details on the policy impacts been given?*
  - *1 point: Have details on policy success been given?*
- **2 points:** Has the document been peer reviewed or independently verified?
  - *1 point: Has the document been peer reviewed by a reputable expert?*
  - *1 point: Has the document been peer reviewed by two or more reputable experts?*
- **2 points:** Are there statements of copyright, regulatory compliance, and possible conflicts of interest present?
  - *1 point: Does the document give statements of copyright, regulatory compliance, or possible conflicts of interest?*
  - *1 point: Does the document acknowledge resource contributions from people or institutions?*
- **1 point:** Does the author/publishing organisation have a track record in the area?
- **1 point:** Where percentages are given, are the totals given?

This will involve skimming the documents and scoring them – they must obtain half of the total number of points (i.e. 7 out of 14) to be included in the final selection. The points are weighted to reflect their importance. Thus, in addition, at least one point must come from either the 'implementation' or 'evaluation' questions, and at least one point must come from another part of the scale. Record the final number of documents in the spreadsheet, but also state why a document was included or excluded (relate to the inclusion/exclusion criteria or study quality assessment)

5. Read the final, included documents and draw out (where stated) the following policy background information, policy impacts (quantitative), and policy mechanisms (qualitative), and input them into the spreadsheet

Background Classification	
Author(s)	
Institution(s)	
Full reference	
Policy year(s)	
Policy category	
Location (country/state)	
Policy implementer(s)	
Sector(s)	
Number of participants	
Policy aims	
Policy methods	
Policy key conclusions	

Policy Impacts	
Overall energy savings	
Carbon savings	
Peak load reductions	
Policy costs to utilities	
Policy benefits to utilities	
Policy costs to government	
Policy benefits to government	
Policy costs to consumers	
Policy benefits to consumers	
Other policy comments	

Policy Mechanisms	
Policy choice details	
Policy design details	
Policy coverage details	
Policy implementation details	
Policy evaluation details	
Policy success details	

By this stage it is likely that there will only be a few papers to look at. Where data and information are not available in the documents, write '*Not stated'* in the relevant cell. With the qualitative aspects, write a few statements regarding what you think is important or interesting to note about each of the areas in the table above, but also what is well covered, what is not discussed in much depth, what is not written, and potential biases

6. Once complete we will compare our findings – both the process of conducting the review and the results

- **Spreadsheet:**
  - Everything is clearly shown in the spreadsheet about what to do
  - 4 tabs for each paper have been created in the template – delete or create tabs (copy and pasting the template) to reflect the number of documents included
  - Roughly speaking, filtering takes about 30-60 minutes and each of the included papers take about 30-60 minutes to extract information from

Appendix Figure 1: instructions document for the systematic review pilot tests

UCL ENERGY INSTITUTE

Dear ‘x’,

Thank you for agreeing to participate in this PhD research, which is funded by the UK Engineering and Physical Sciences Research Council (EPSRC) and supervised by Professor Paul Ekins and Dr Mark Barrett at University College London (UCL).

Project overview
The project looks at how and why demand-side management (DSM) policies succeed or fail and aims to answer the following research questions:

1. What DSM policies have been implemented around the world with high quality documented evaluations?
2. How and why do DSM policies succeed or fail, and what policies have been successful?

To answer these research questions a systematic review of documents evaluating DSM policies is employed by consulting academic, industrial and policy databases. The findings from the systematic review will be enriched by the insights of key experts in the DSM field.

Expectations of the meeting
The method used to capture your expert insights is Multi-Criteria Decision-Making (MCDM) analysis. You will be invited to fill in a short spreadsheet ranking (from 1-10) different DSM policies based on various criteria, such as carbon savings and peak load reductions. You will then be asked more generally for your thoughts on the following three questions:

1) From your experience, what have been the most successful DSM policies?
2) How and why do you think these policies have been successful?
3) What are the key factors that may result in a DSM policy being unsuccessful?

The meeting will last around 45 minutes (5 minutes to introduce the project, 20 minutes to fill in the spreadsheet, and 20 minutes to discuss the questions). Any information that you provide will be treated confidentially and handled in accordance with the UK Data Protection Act 1998. Reporting is anonymous and you will not be identified by name; if you are quoted, you will only be referred to as an expert from ‘y’. If for any reason you want to withdraw from the study, you may do so at any time. If you have any questions/comments about my research, please do not hesitate to contact me.

Yours sincerely,

Peter Warren
Doctoral Researcher
Energy Institute, University College London, 14 Upper Woburn Place, London, WC1H 0NN, UK
E: peter.warren.10@ucl.ac.uk, T: +447725978045

Appendix Figure 2: confidentiality form for MCDM participants

Appendix Figure 3: decision matrix and questions for the MCDM analysis with DSM policy experts

MCDM Analysis Decision Matrix

<u>Instructions:</u>

1) Please fill out the table below by putting a number from 1-10 in each box (fill across by rows for each policy individually):
    a. *1 – lowest value (e.g. lowest reductions or highest costs)
    b. *10 – highest value (e.g. largest reductions or lowest costs)
    c. *Please ignore the OS (Overall Score) and PR (Policy Ranking) columns
2) On the ‘Questions’ tab, please fill out the four boxes

DSM Policy Type / DSM Policy Criteria	CO_{2e} reduction	Deferred investment in infrastructure	Energy bill savings	Government programme costs	Overall energy savings	Peak load reductions	Utility programme costs	Consumer active engagement	Dealing with variable wind power	Political ease of implementation	Technology innovation & market development	OS	PR
Incentive payment-based DR												0	
Price-based DR												0	
Market transformations												0	
Infrastructure rollouts												0	
Utility obligations												0	
Labelling												0	
Performance standards												0	
Loans and subsidies												0	
Utility business												0	

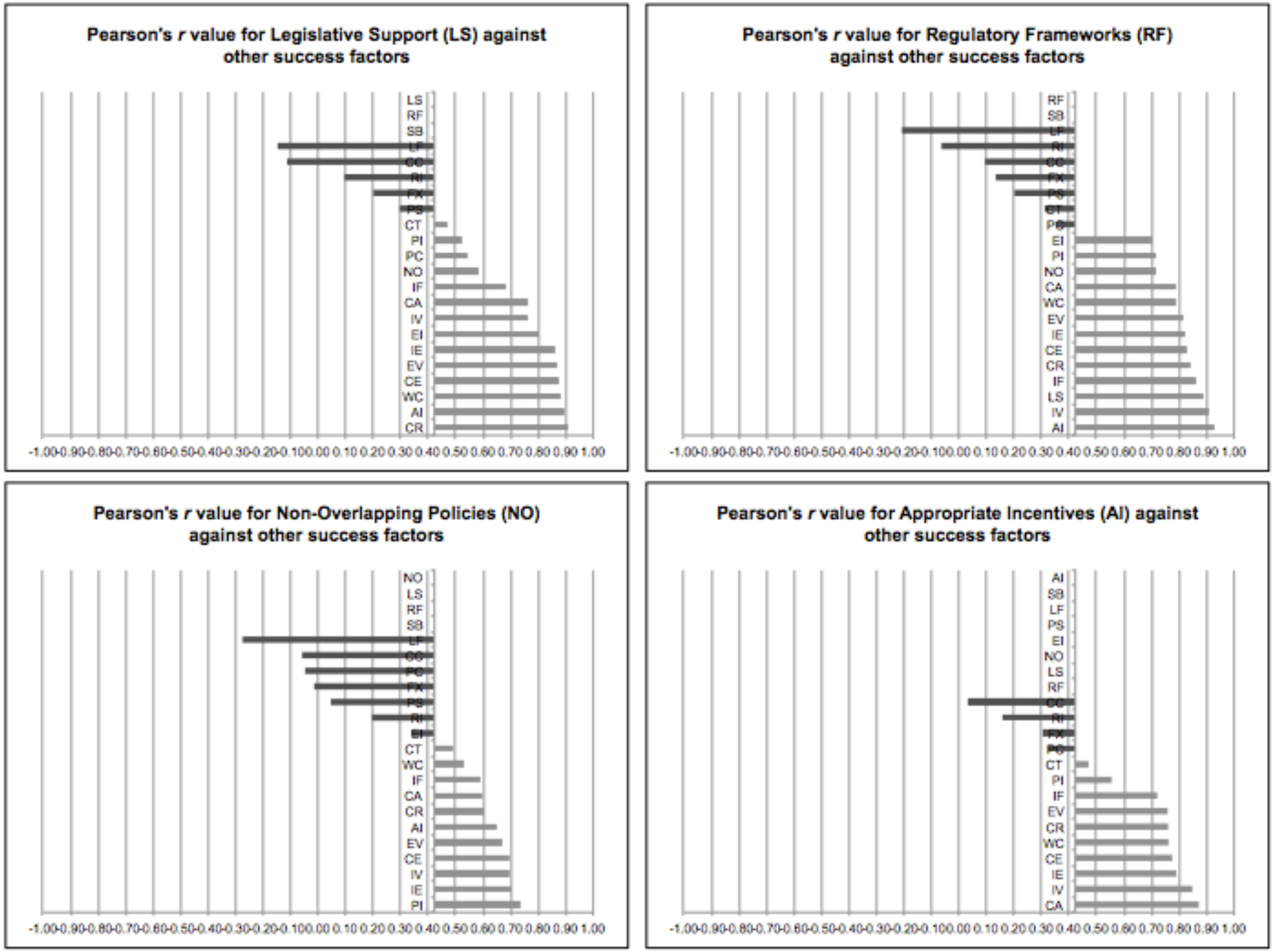
(Appendix Figure 3 continued)

1) From your experience, what have been the most successful DSM policies?

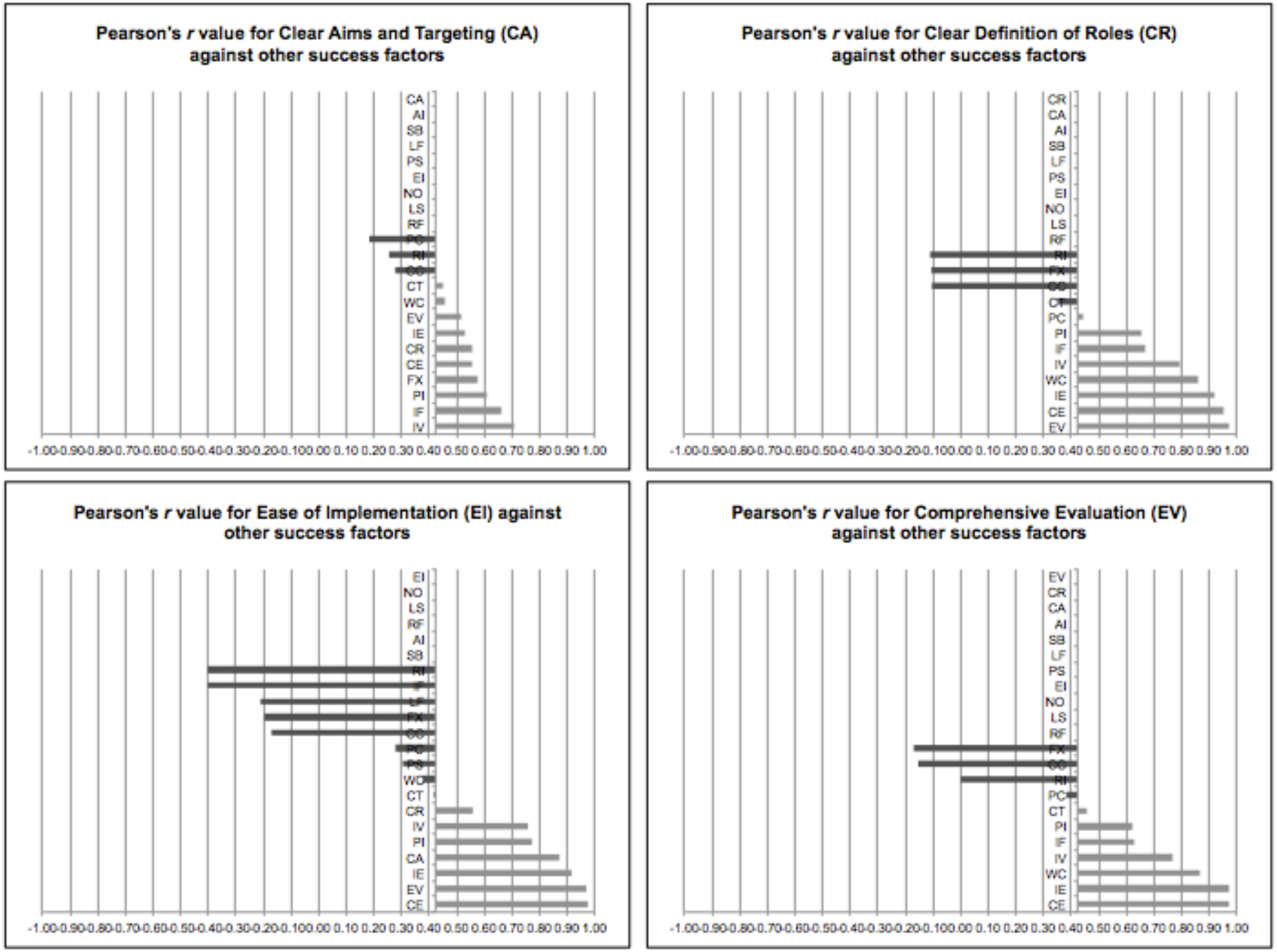
2) How and why do you think these policies have been successful?

3) What are the key factors that may result in a DSM policy being unsuccessful?

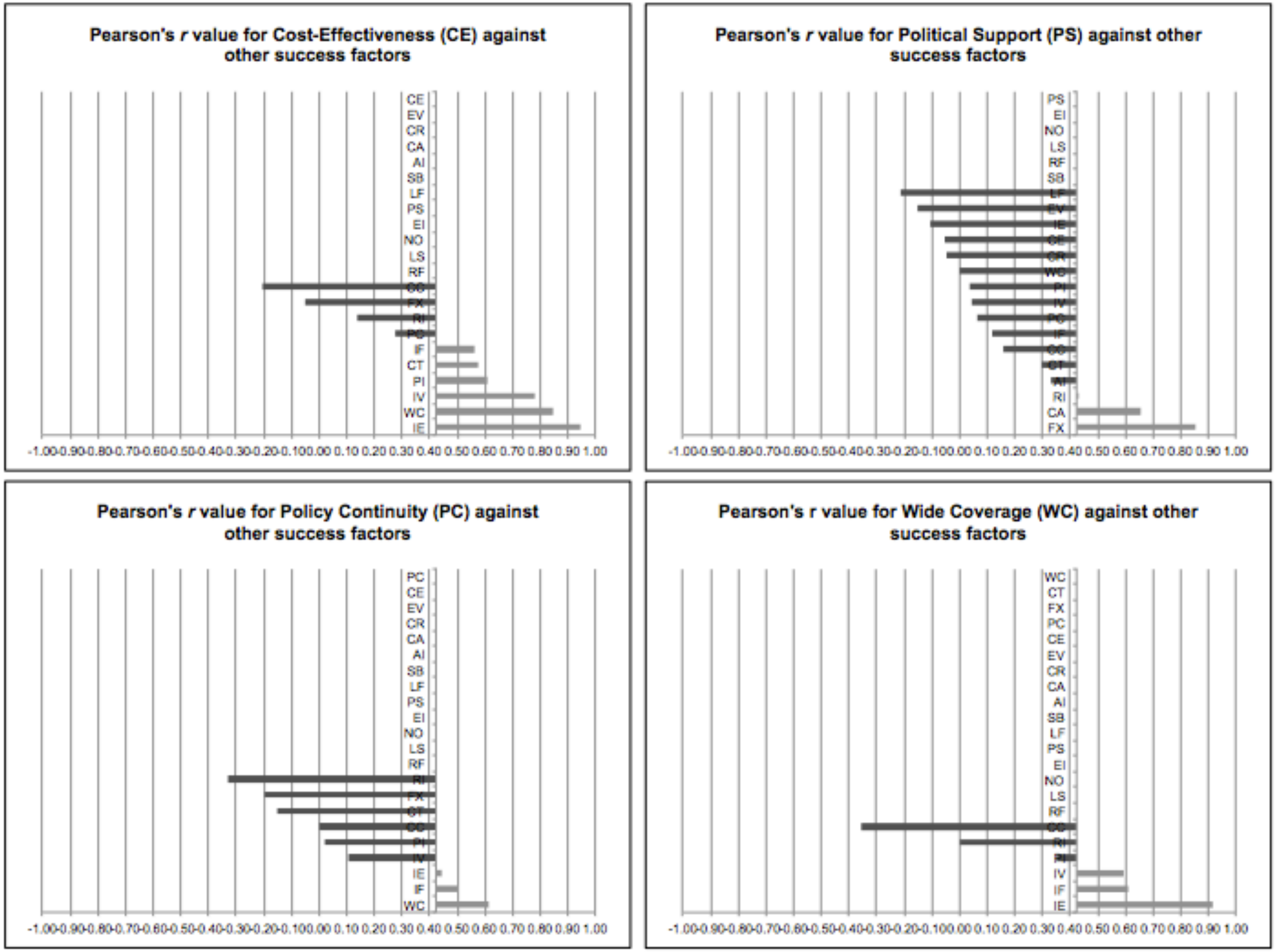
Appendix Figure 4: statistical associations between DSM policy success factors (22 graphs for 22 success factors)



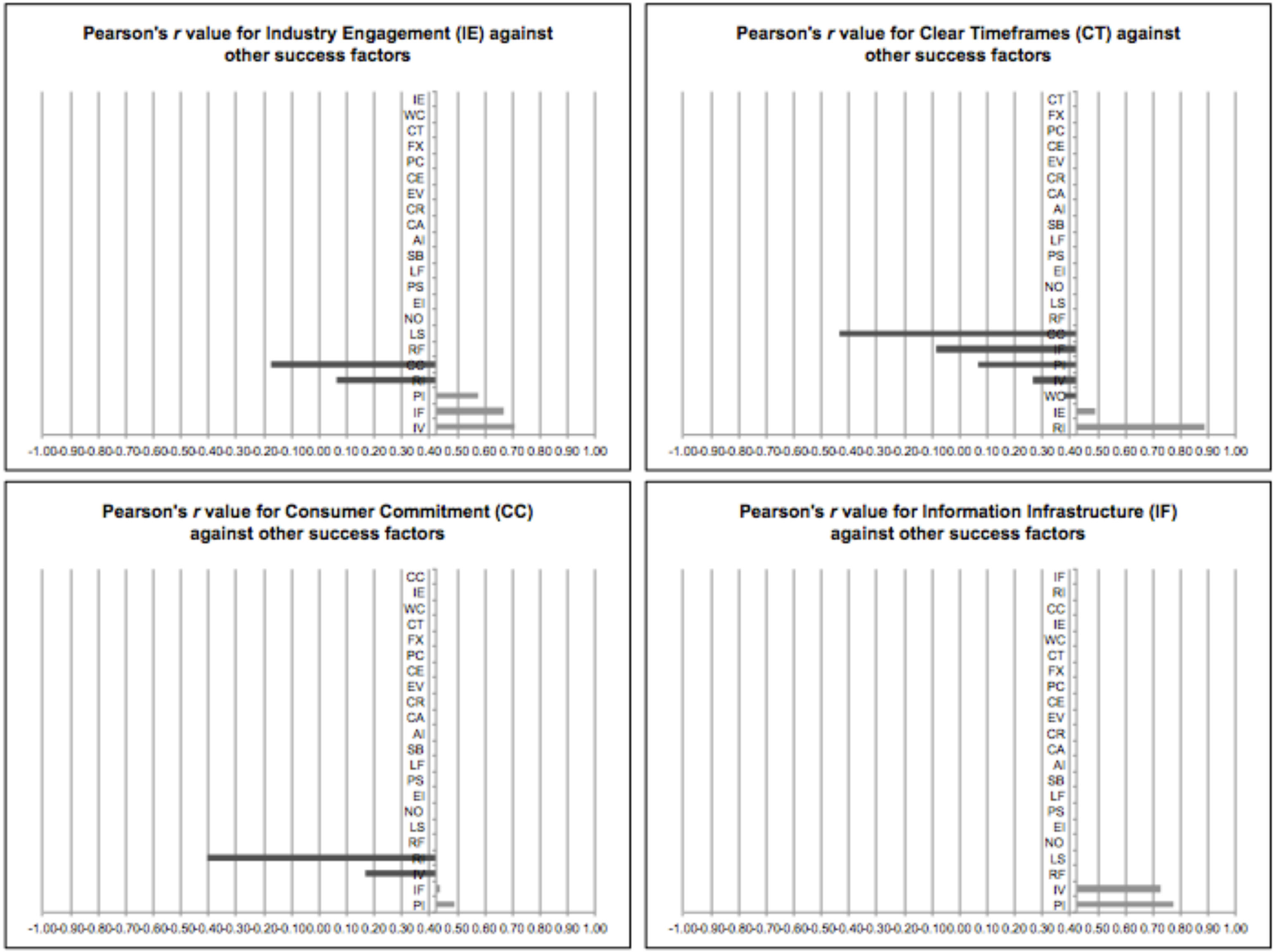
(Appendix Figure 4 continued)



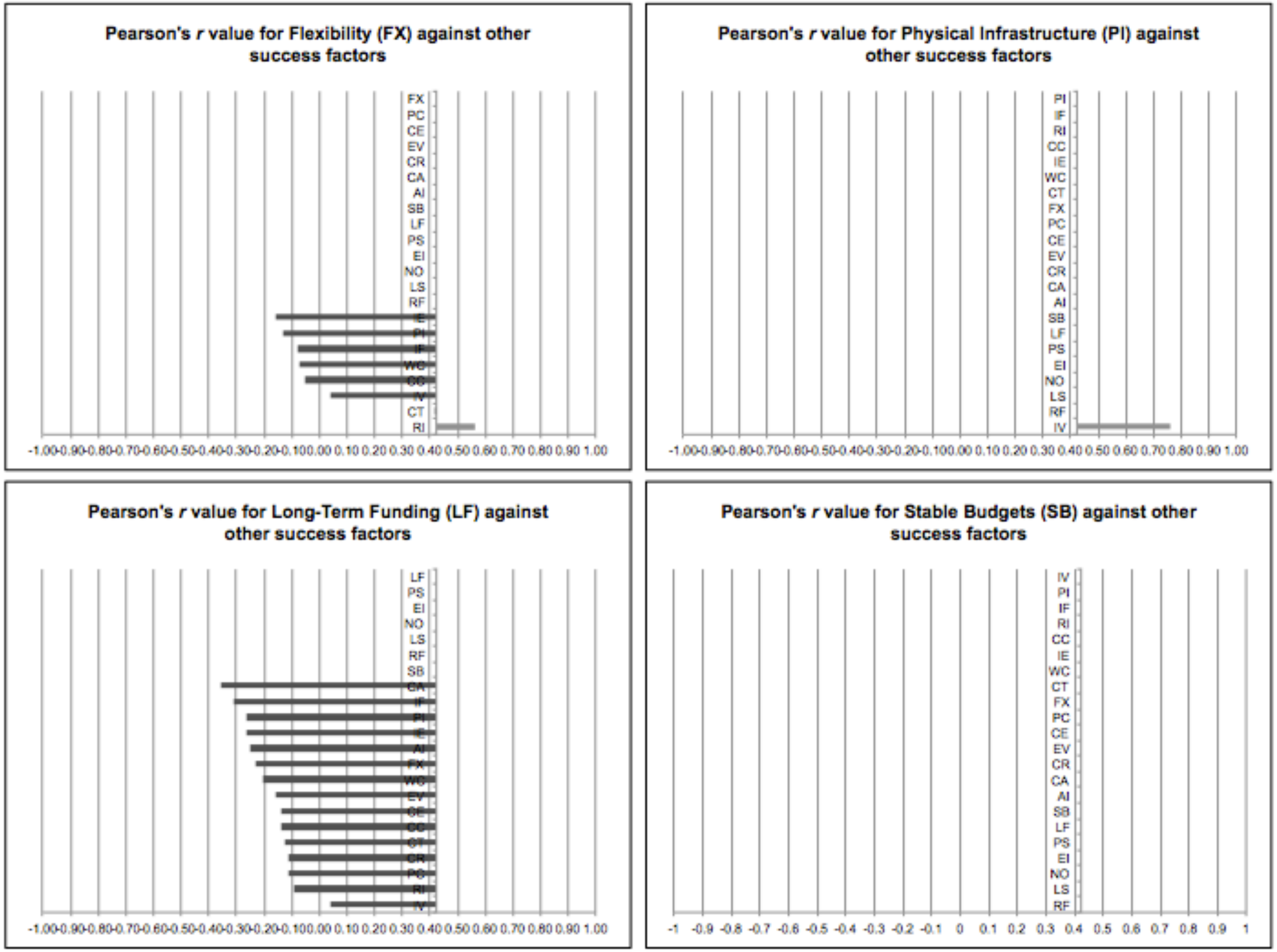
(Appendix Figure 4 continued)



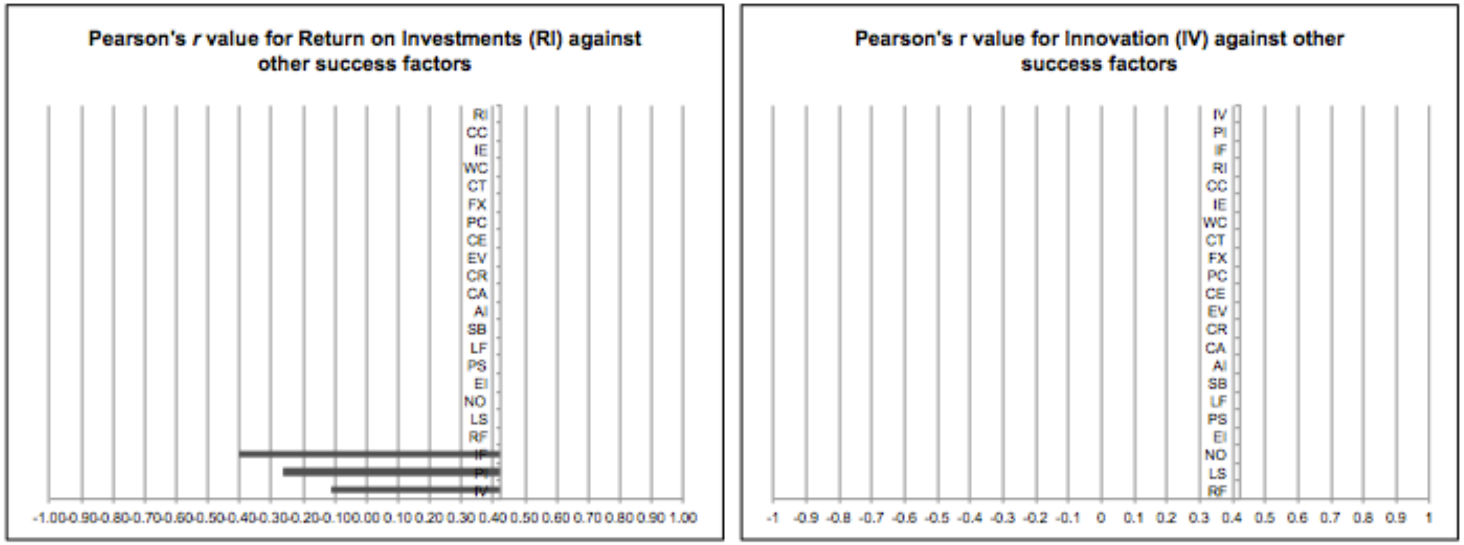
(Appendix Figure 4 continued)



(Appendix Figure 4 continued)

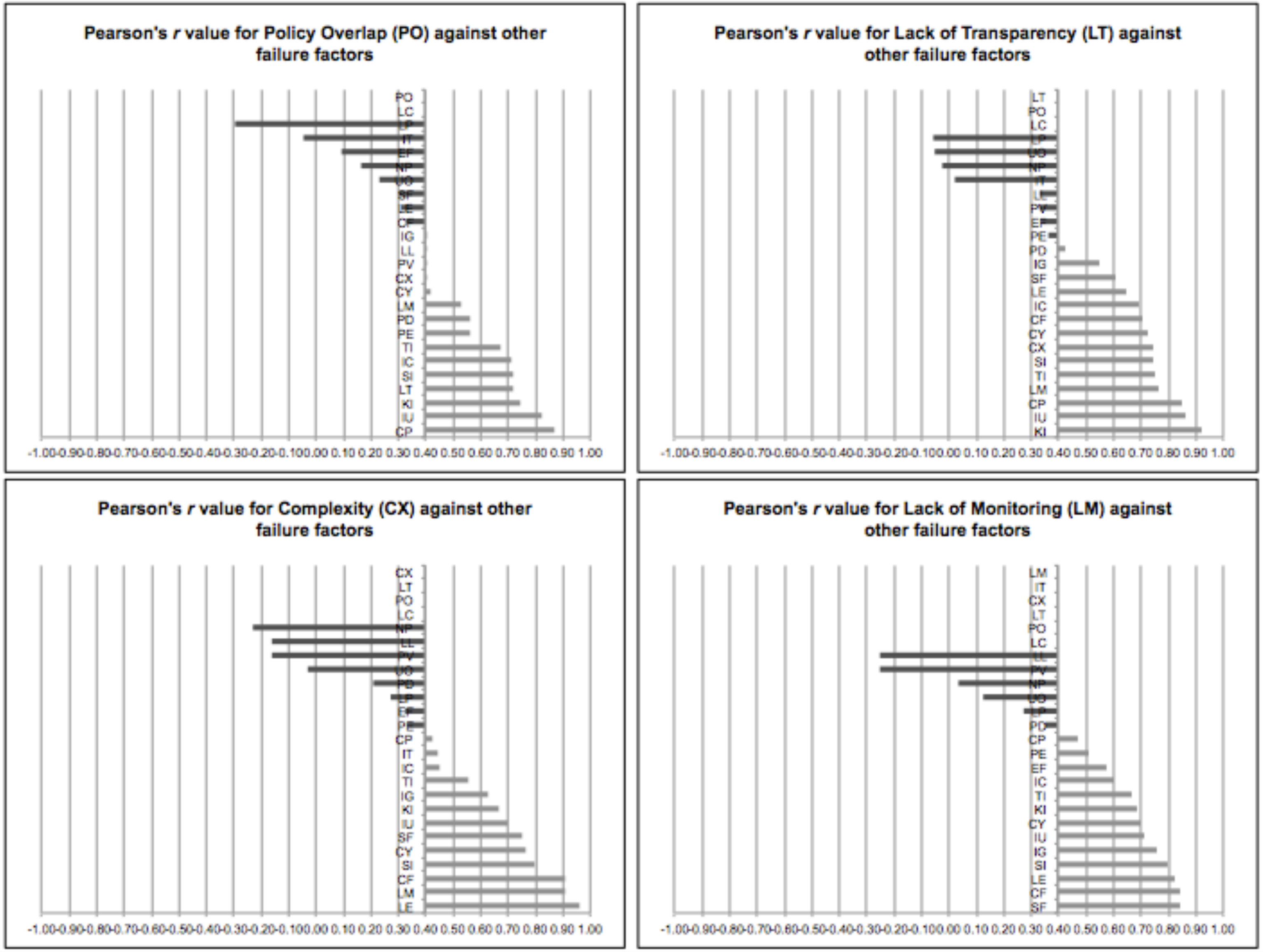


(Appendix Figure 4 continued)

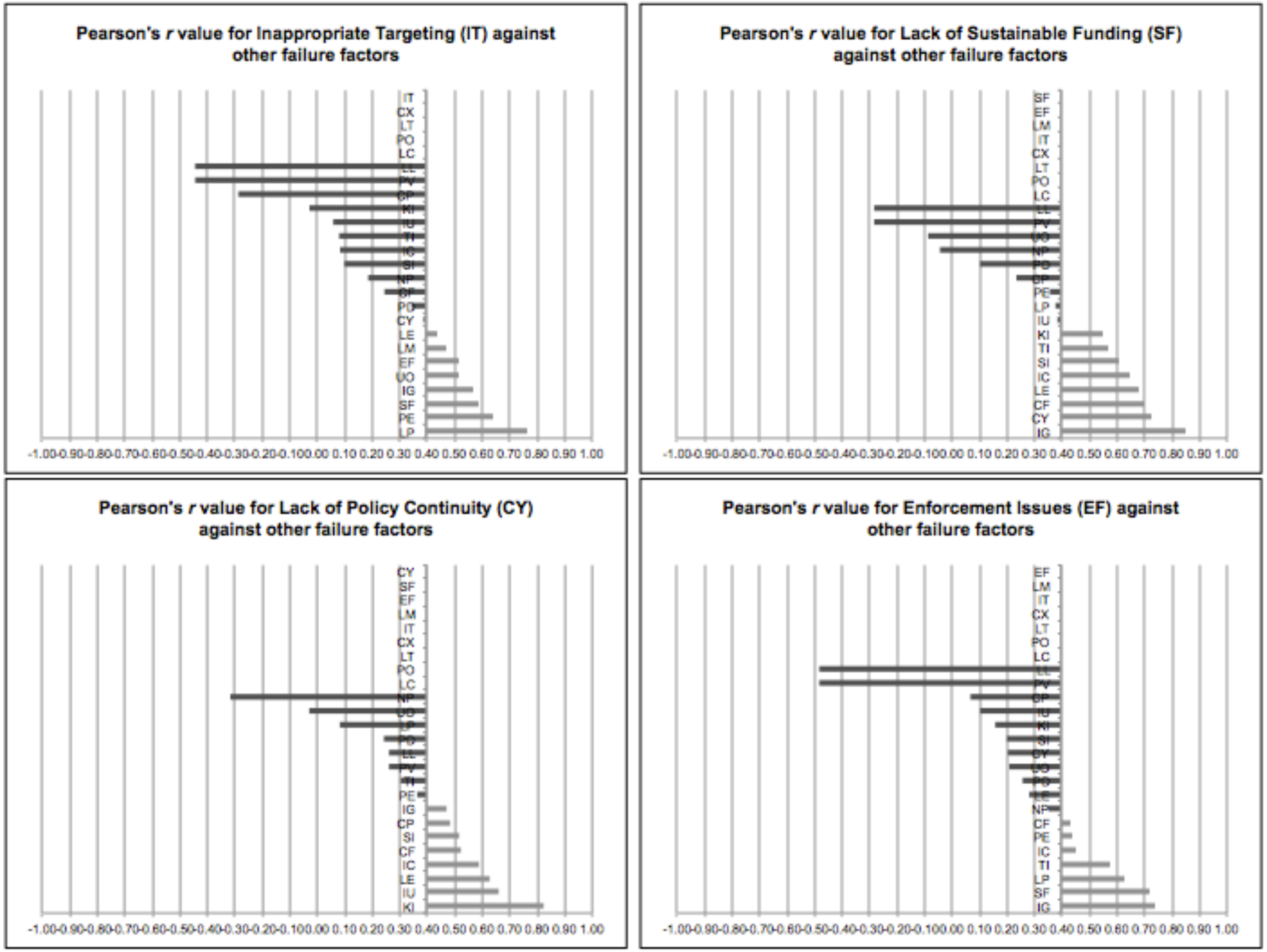


Success Factor	Acronym	Success Factor	Acronym	Success Factor	Acronym
Regulatory frameworks	RF	Clear aims and targeting	CA	Wide coverage	WC
Legislative support	LS	Clear definition of roles	CR	Industry engagement	IE
Non-overlapping policies	NO	Comprehensive evaluation	EV	Consumer commitment	CC
Ease of implementation	EI	Cost-effectiveness	CE	Return on investments	RI
Political support	PS	Policy continuity	PC	Information infrastructure	IF
Long-term funding	LF	Flexibility	FX	Physical infrastructure	PI
Stable budgets	SB	Clear timeframes	CT	Innovation	IV
Appropriate incentives	AI				

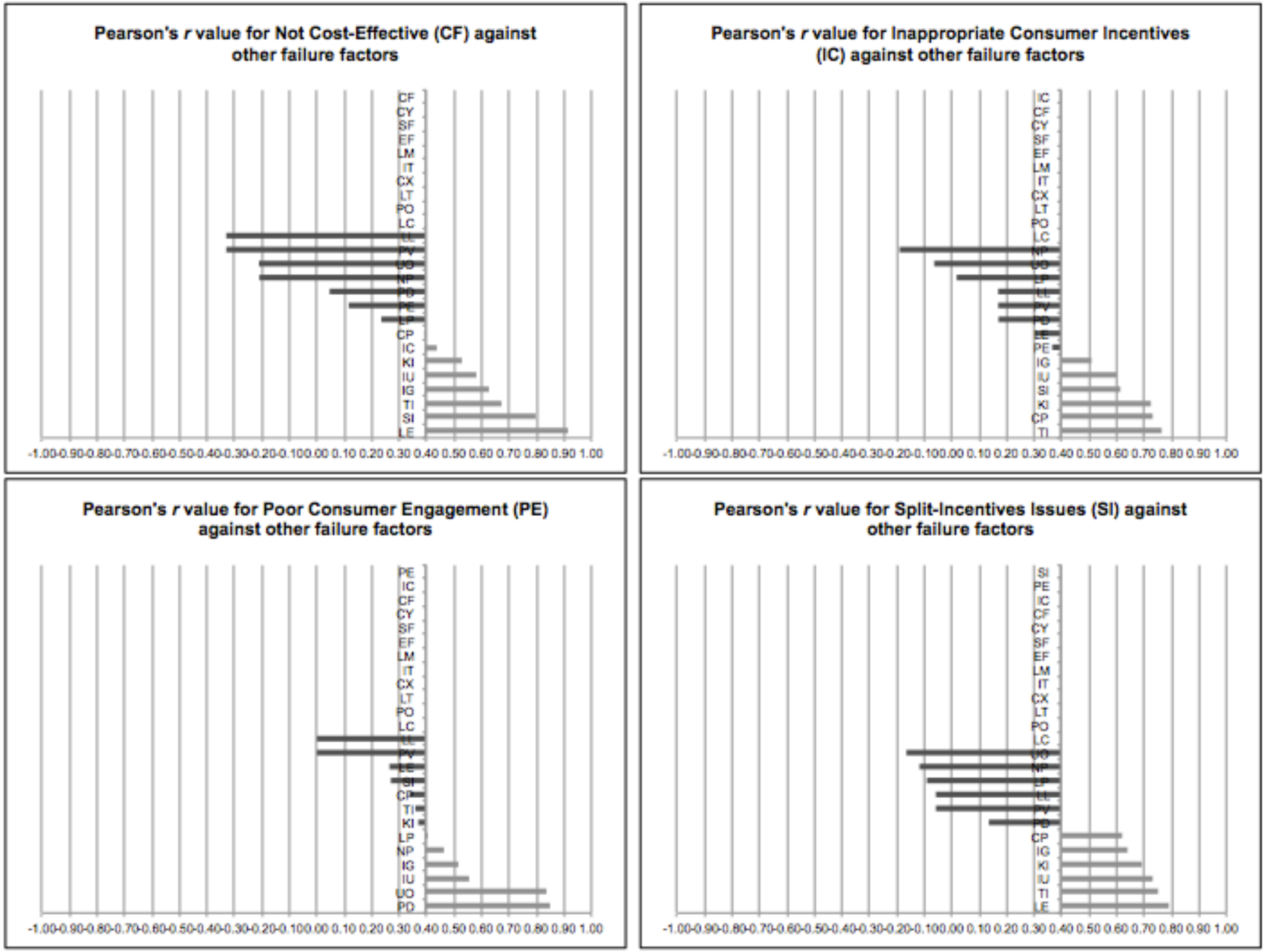
Appendix Figure 5: statistical associations between DSM policy failure factors (25 graphs for 25 failure factors)



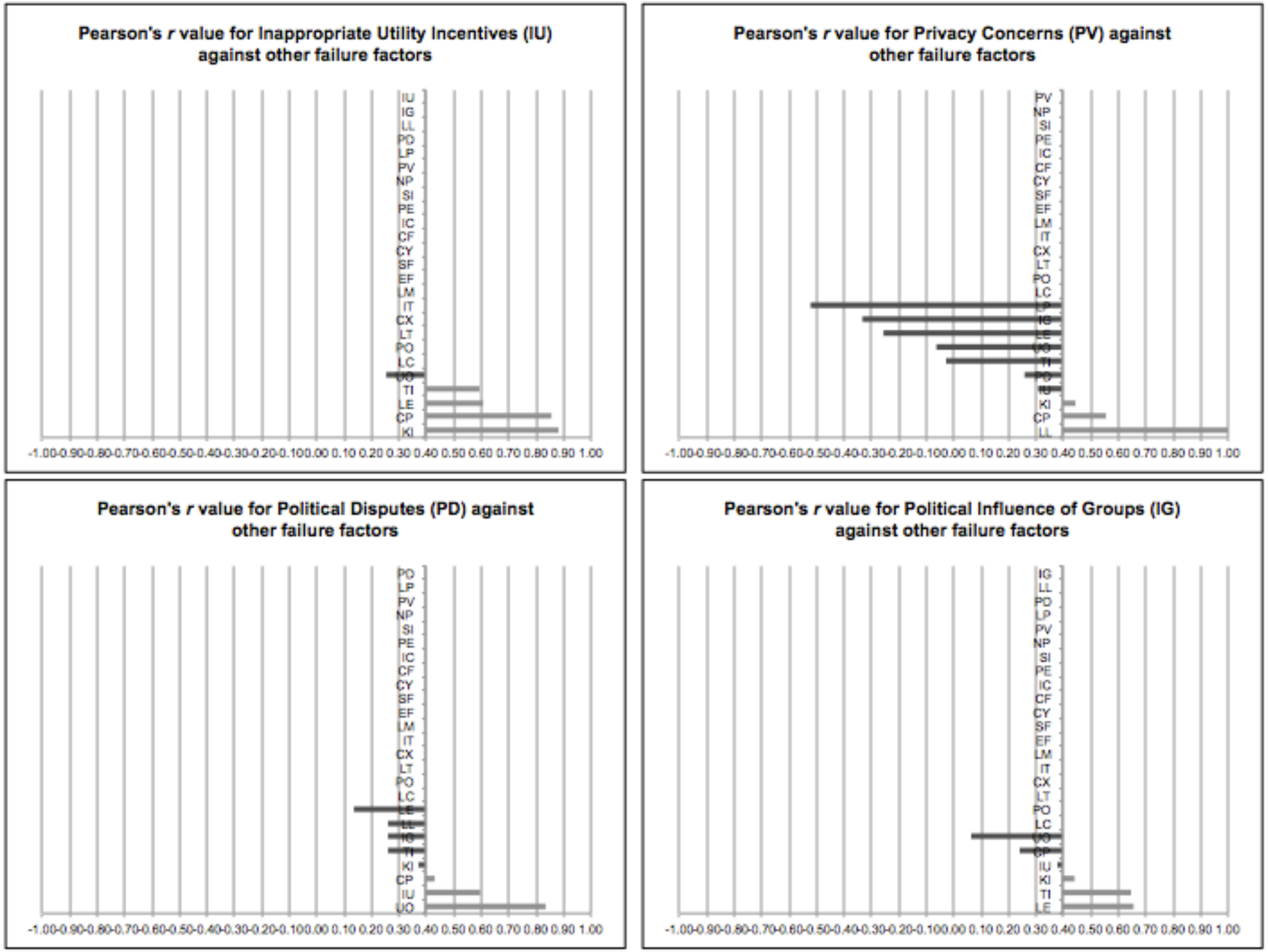
(Appendix Figure 5 continued)



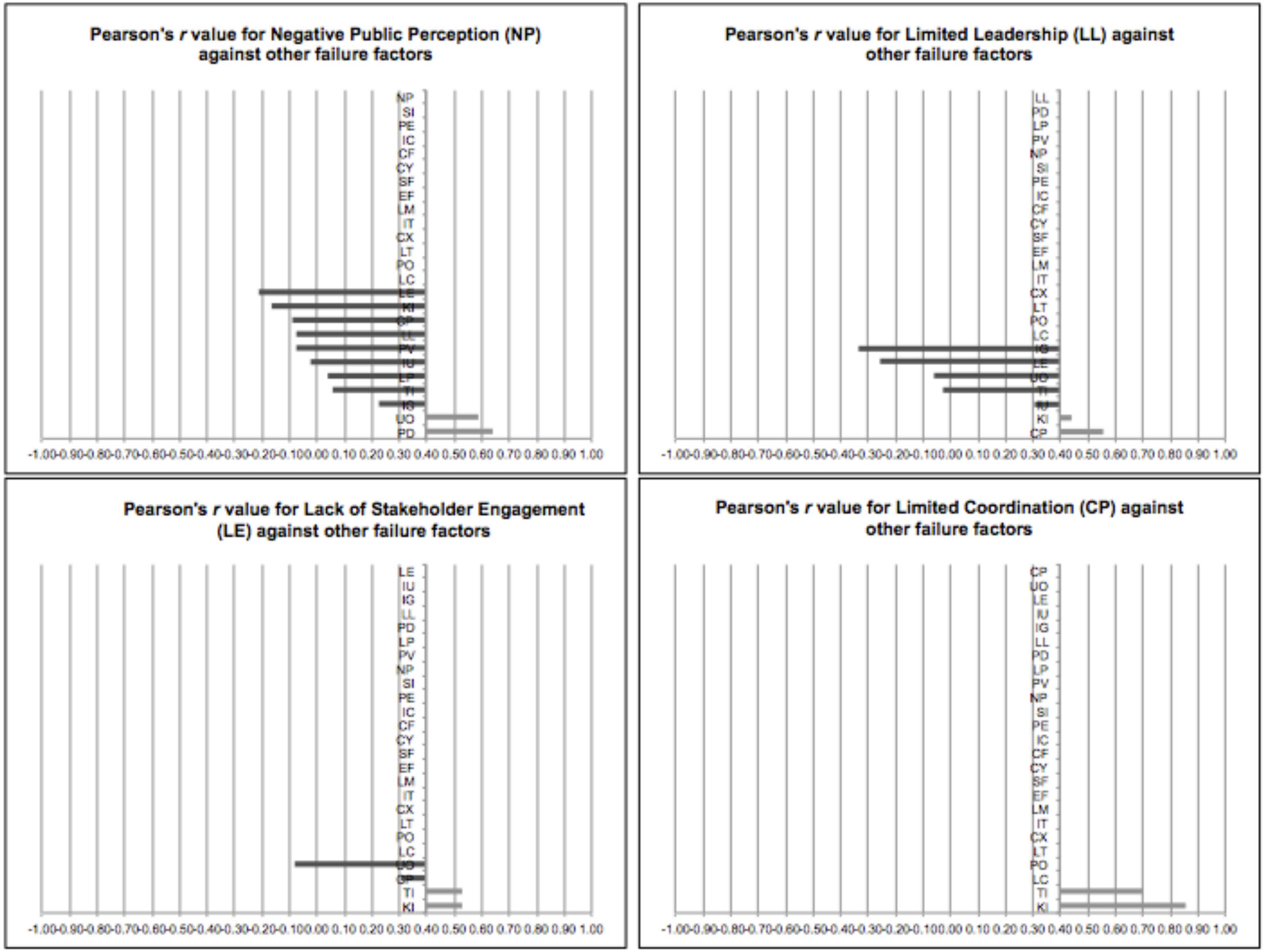
(Appendix Figure 5 continued)



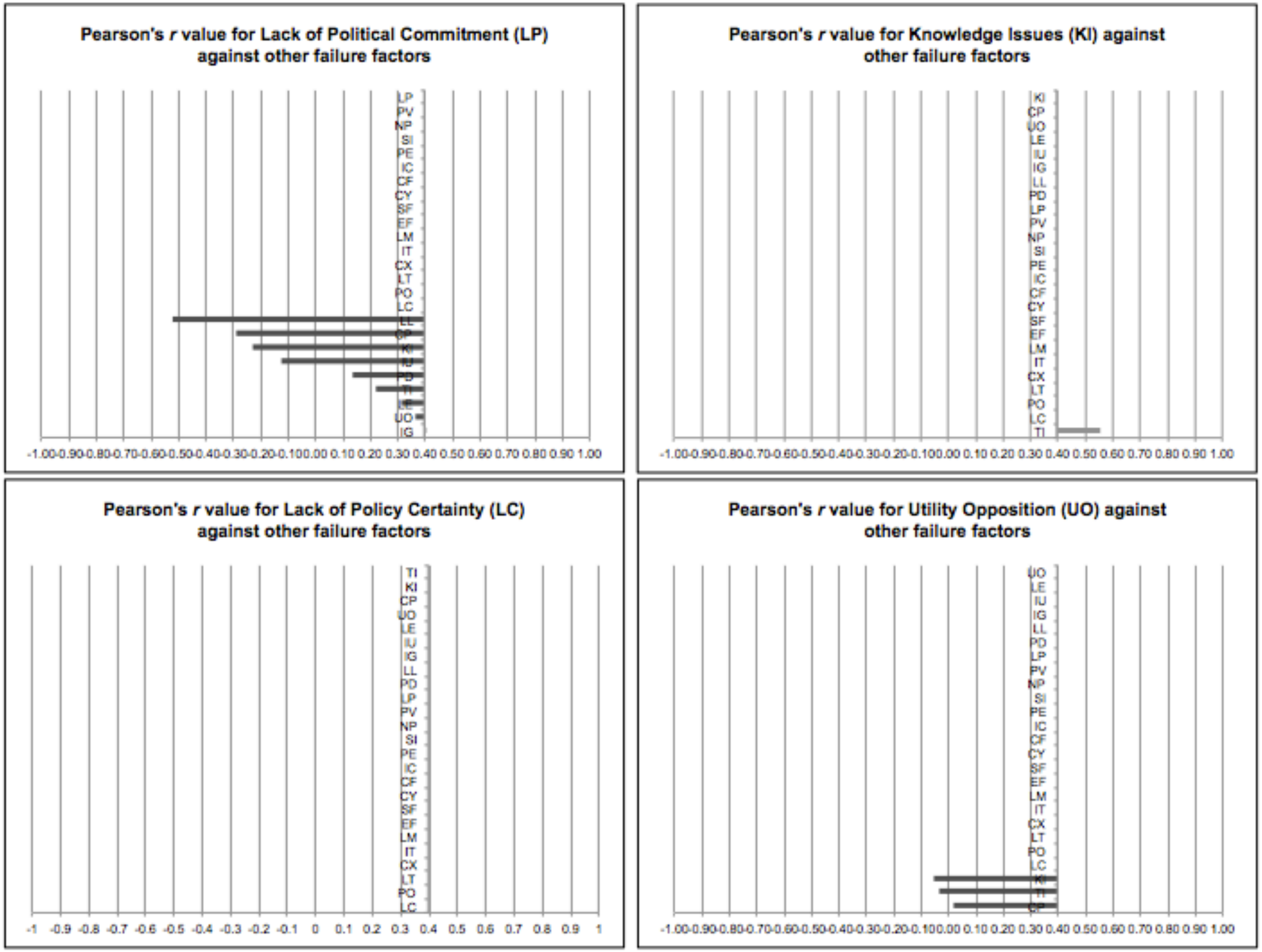
(Appendix Figure 5 continued)



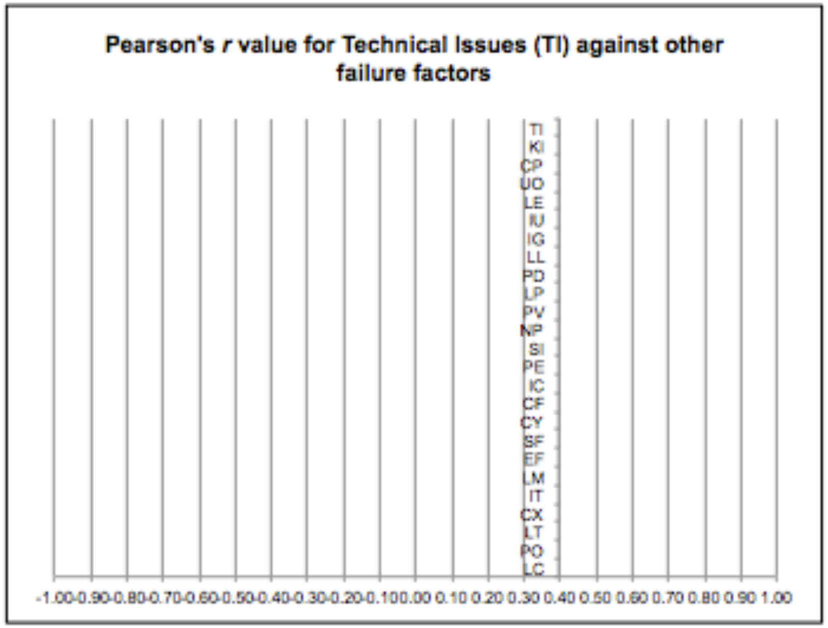
(Appendix Figure 5 continued)



(Appendix Figure 5 continued)



(Appendix Figure 5 continued)



Failure Factor	Acronym	Failure Factor	Acronym	Failure Factor	Acronym
Lack of policy certainty	LC	Not cost-effective	CF	Limited leadership	LL
Policy overlap	PO	Inappropriate consumer incentives	IC	Political influence of groups	IG
Lack of transparency	LT	Poor consumer engagement	PE	Inappropriate utility incentives	IU
Complexity	CX	Split-incentives issues	SI	Lack of stakeholder engagement	LE
Inappropriate targeting	IT	Negative public perception	NP	Utility opposition	UO
Lack of monitoring	LM	Privacy concerns	PV	Limited coordination	CP
Enforcement issues	EF	Lack of political commitment	LP	Knowledge issues	KI
Lack of sustainable funding	SF	Political disputes	PD	Technical issues	TI
Lack of policy continuity	CY				

Country/State	Successful Policies	Number of Successful Policies
USA (California)	IPBDR, PBDR, MT, IR, PS, L&S, UBM, R&D, IC, UBM/MT	**10**
China	IPBDR, PS, VP, PS/LB, IC/L&S, PBDR, IC, LB, L&S, R&D	**10**
UK	IPBDR, PBDR, IR, UO, PS, L&S, UBM, R&D, IC	**9**
USA	IPBDR, PBDR, UO, PS, L&S, UBM, R&D, UBM/MT, IC/L&S	**9**
Denmark	UO, LB, PS, L&S, R&D, IC, VP	**7**
Thailand	MT, LB, L&S, IC	**4**
USA (New York)	IPBDR, L&S, UBM, UBM/MT	**4**
USA (Vermont)	PBDR, UO, PS, UBM	**4**
USA (state-level)	UO, PS, UBM	**3**
Germany	L&S, IC, IC/L&S	**3**
USA (PJM region)	PBDR, IPBDR/PBDR	**2**
France	PBDR, UO	**2**
USA (Pacific Northwest region)	UBM/MT, PS/IC	**2**
USA (Massachusetts)	UBM/MT, IC/L&S	**2**
European Union (EU)	UO, PS	**2**
Australia	IR, PS	**2**
USA (NYISO region)	IPBDR/PBDR	**1**
China (Hebei)	UBM	**1**
China (Fujian)	UBM	**1**
Belgium (Flanders)	UO	**1**
Italy	UO	**1**
Japan	UO	**1**
Brazil	UO	**1**
Estonia	L&S	**1**
USA (ISO-NE region)	IPBDR/PBDR	**1**
USA (Illinois)	IC/L&S	**1**
USA (Florida)	IPBDR	**1**
China (Jiangsu)	IPBDR/PBDR	**1**
China (Beijing)	IPBDR/PBDR	**1**
India (Orissa)	L&S	**1**
Australia (New South Wales)	UO	**1**
Australia (ACT)	UO	**1**
Australia (South Australia)	UO	**1**
Australia (Victoria)	UO	**1**
Philippines	PS/LB/IC	**1**
USA (Ohio)	UBM	**1**
Spain	IPBDR	**1**
USA (Wisconsin)	IC/L&S	**1**
Canada	UO	**1**
South Korea	IC	**1**
Sweden	MT	**1**

Appendix Figure 6: countries/states that have successfully implemented various DSM policies

Country/State	Unsuccessful Policies	Number of Unsuccessful Policies
European Union (EU)	IPBDR, PBDR, IR, LB, L&S, IC, PS/LB	**7**
India	IPBDR, LB, PS, L&S, R&D, IC, VP/L&S	**7**
Mexico	IPBDR, LB, PS, L&S, R&D, IC, IC/L&S	**7**
USA	MT, IR, LB, IC, VP, IPBDR/PBDR	**6**
Australia	IPBDR, MT, UO, LB, IC	**5**
Netherlands	UO, LB, PS, L&S, IC	**5**
South Korea	IPBDR, PBDR, LB, L&S, VP	**5**
South Africa	IPBDR, PBDR, UO, L&S	**4**
UK	MT, LB, L&S, VP/L&S	**4**
Canada	LB, PS, L&S	**3**
Croatia	LB, PS, IC	**3**
Japan	MT, IR, LB	**3**
USA (state-level)	MT, L&S, UBM/MT	**3**
USA (Wisconsin)	UBM, R&D, PS/IC	**3**
Canada (British Columbia)	IPBDR, PBDR, IR	**3**
Canada (Ontario)	UO, L&S, UBM	**3**
China	UBM, IPBDR/PBDR	**2**
France	L&S, R&D	**2**
Germany	PS, VP	**2**
New Zealand	IPBDR, IC/L&S	**2**
Pakistan	IC, PS/LB	**2**
Sweden	LB, IC	**2**
USA (California)	IC/L+S/MT, PS/LB/UO/L&S	**2**
USA (Ohio)	IPBDR, IC/L&S	**2**
USA (Oregon)	UO, L&S	**2**
Belgium	PS/IC	**1**
Denmark	UBM	**1**
Indonesia	IC	**1**
Ireland	IC	**1**
Italy	PS/LB/UO/L+S	**1**
Kenya	L&S	**1**
Philippines	R&D	**1**
Spain	MT	**1**
Sri Lanka	IC/L+S	**1**
USA (Illinois)	IC/L+S	**1**
USA (New York)	PS	**1**
USA (Maine)	IC/L+S	**1**
USA (New Hampshire)	IC/L+S	**1**
USA (Michigan)	UBM	**1**
USA (Texas)	PBDR	**1**
China (Jiangsu)	PS/LB	**1**
China (Shanghai)	PS/LB	**1**
China (Beijing)	PS/LB	**1**
China (Guangzhou)	PS/LB	**1**
China (Hefei)	PS/LB	**1**
China (Shandong)	PS/LB	**1**
China (Sichuan)	PS/LB	**1**

Appendix Figure 7: countries/states that have unsuccessfully implemented various DSM policies